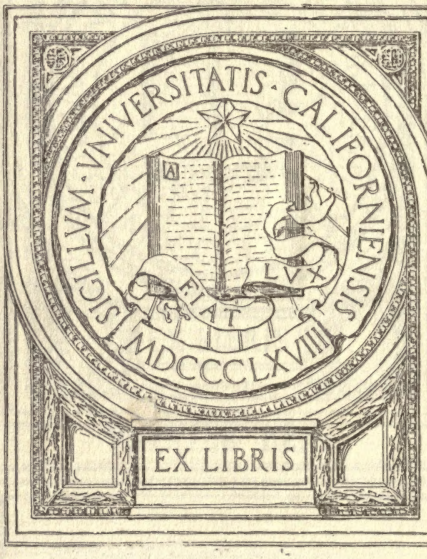
GIFT OF
Prof. E. P. Lewis



E. P. Lewis

A MANUAL OF PHYSICS.

BEING

AN INTRODUCTION TO THE STUDY OF PHYSICAL SCIENCE.

Designed for the Use of University Students.

BY

WILLIAM PEDDIE, D.Sc., F.R.S.E.,

ASSISTANT TO THE PROFESSOR OF NATURAL PHILOSOPHY IN THE UNIVERSITY OF EDINBURGH.

NEW YORK: G. P. PUTNAM'S SONS.
LONDON: BAILLIÈRE, TINDALL & COX.
1892.

QC21
P44

Gift of Prof E. P. Lewis
PHYSICS DEPT.

PREFACE.

THE best advice which can be given to a student of physics regarding the books which he should read is to use separate works written on the various branches of the subject by the leading physicists of the day. Yet this advice has one evident disadvantage. The student who follows it may not get so complete a view of the essential unity and interdependence of the various branches of his subject as it is desirable that he should. And, besides this, there is no doubt that a small volume which gives, as far as is possible, a review of the elements of the whole subject, is a desideratum to the student while in attendance on University classes. I have undertaken the writing of this work in the hope that it may to some extent meet that want.

I have throughout endeavoured to bring into prominence the necessity for, and the value of, scientific hypotheses—a matter regarding which very hazy notions are only too common.

It has also been my aim to make the treatment of the mathematical portions of the text as simple as possible. In this connection I have not adopted the process which has recently been termed 'calculus-dodging,' for the reason that the elementary methods of the calculus are more simple, certainly are more natural, than the methods by which they are usually supplanted.

At the same time it may be well to remark that any student, who desires to do so, may simply assume the results of the mathematical portions, and use the remainder (which is much the larger part) of the text in his study of experimental physics.

In writing a text-book on general physics it is impossible, if justice is to be done to the subject, to avoid borrowing methods from the writings of the masters. In this respect I have to acknowledge my indebtedness to the works of v. Helmholtz, Clerk-

357598

Maxwell, Thomson, Tait, and others. This is perhaps most evident in Chapter XXVI., where I have made use of the very simple form in which Tait has presented the analytical treatment of the theory of Thermodynamics; and in Chapter XI., where I have adopted his mode of discussing the compressibility and rigidity of solids. This apart, I have endeavoured, whether successfully or not, to present the various subjects in as fresh a manner as I could.

My indebtedness to Professor Tait has also been very great in the matter of criticism, which he kindly afforded me on various points while the book was passing through the press. I have also to acknowledge with thanks the kindness of Mr. J. B. Clark, M.A., F.R.S.E., Physical Master in Heriot's Hospital, Edinburgh, in reading the work both in manuscript and in proof. His careful revision has resulted in the elimination of a number of defects which had escaped my notice.

WILLIAM PEDDIE

EDINBURGH UNIVERSITY,
November, 1891.

CONTENTS.

CHAPTER I.

INTRODUCTORY: THE PHYSICAL UNIVERSE.

PAGE

Physical science.—Test of reality: conservation.—Matter: its conservation: its inertia.—Energy: its transformation: its conservation: its degradation . . . . 1—8

CHAPTER II.

THE METHODS OF PHYSICAL SCIENCE.

Observation and experiment.—Cause and effect.—Hypothesis and theory.—Crucial test.—Argument from analogy—Principle of stable equilibrium.—Errors of observation.—Instrumental errors.—Empirical laws.—The graphical method . . . . . 9—17

CHAPTER III.

THE THEORY OF CONTOURS, AND ITS PHYSICAL APPLICATIONS.

Dimensions.—Contour points of plane and tortuous curves.—Examples of their applications in physics.—Contour lines.—Special physical applications.—Use of trilinear co-ordinates . . . . . 18—32

CHAPTER IV.

VARYING QUANTITIES.

Dimensions of physical quantities.—Determination of rates of variation.—The inverse problem . . . 33—47

CHAPTER V.

MOTION.

PAGE

Position.—Displacement.—Speed and velocity.—Acceleration of speed.—Curvature.—Acceleration of velocity.—Acceleration in, and perpendicular to, the direction of motion.—Average speed and velocity.—Resolution and composition of velocities and accelerations.—Motion of projectiles.—The hodograph.—Moments.—Acceleration perpendicular to the radius-vector.—Simple harmonic motion.—Composition of simple harmonic motions.—Wave motion along a line.—Rotation.—Alteration of co-ordinates because of rotation.—Uniplanar motion of a rigid body.—Motion of a rigid body in space.—Composition of angular velocities.—Displacement of the parts of a non-rigid body.—Strain.—Homogeneous strain.—Shear.—Non-homogeneous strain.—Motion of fluids.—Fluid circulation.—Vortex motion 48—78

CHAPTER VI.

MATTER IN MOTION.

Force.—The laws of motion.—The first and second laws.—Examples.—Dynamical similarity.—The third law.—Centre of inertia.—Moment of a force and of inertia.—Rotational equilibrium.—Propagation of motion through a non-rigid solid.—Motion of a perfect fluid.—Equilibrium of a fluid.—Propagation of surface-waves in liquid . 79—100

CHAPTER VII.

PROPERTIES OF MATTER.

Definitions of matter.—States of matter.—General properties.—Special properties.—Specific properties.—Molecules and atoms.—Molecular forces . . . . 101—106

CHAPTER VIII.

GRAVITATION.

Weight and mass.—Kepler's laws with Newton's deductions.—The law of gravitation.—Newton's theorems regarding spherical shells.—Mean density of the earth.—Hypotheses

PAGE

framed to explain gravitation.—The nebular hypothesis.—Potential.—Gravitational potential.—Equipotential surfaces.—Lines and tubes of force.—Special theorems 107—124

CHAPTER IX.

PROPERTIES OF GASES.

Compressibility.—Boyle's law.—Compressibility of a perfect gas.—Deviations from Boyle's law.—Compression of vapours.—Elasticity.—Viscosity.—Diffusion.—Effusion.—Transpiration . . . . . 125—133

CHAPTER X.

PROPERTIES OF LIQUIDS.

Compressibility.—Elasticity.—Viscosity and viscidity.—Diffusion.—Osmose.—Dialysis.—Cohesion.—Capillarity.—Surface-tension . . . . . 134—144

CHAPTER XI.

PROPERTIES OF SOLIDS.

Compressibility and rigidity.—Elasticity.—Viscosity.—Cohesion—Tenacity . . . . . 145—153

CHAPTER XII.

THE CONSTITUTION OF MATTER.

The hard atoms of Lucretius.—Boscovich's centres of force.—Vortex atoms.—Heterogeneity of matter.—Structure of crystals.—Molecular nature of matter.—Size of molecules and the range of molecular forces . . . 154—162

CHAPTER XIII.

THE KINETIC THEORY OF MATTER.

Kinetic theory of gases.—Gaseous pressure.—Boyle's law.—Avogadro's and Charles' laws.—Diffusion.—Thermal conductivity.—Viscosity.—Evaporation.—Dissociation, etc.—Difficulties of the theory.—General kinetic theory of matter . . . . . 163—170

CHAPTER XIV.

SOUND.

PAGE

Propagation. — Intensity. — Reflection. — Refraction. — Diffraction. — Interference. — Pitch. — Musical intervals. — Vibrations of rods. — Vibrations of plates. — Vibrations of strings. — Vibrations of air-columns. — Speed of sound in a gas — Partial tones. — Resonance. — Quality. — Beats. — Consonance and dissonance. — Combination tones. — Dissonance of pure tones . . . . . 171—191

CHAPTER XV.

LIGHT: INTENSITY, SPEED, THEORIES.

Rectilinear propagation. — Intensity. — Speed. — Theories. — Colour . . . . . 192—197

CHAPTER XVI.

LIGHT: REFLECTION, REFRACTION, DISPERSION.

Laws of reflection. — Reflection from plane surfaces. — Reflection from curved surfaces. — Caustics. — Focal lines. — Theoretical deductions of the laws of reflection. — Laws of refraction. — Refraction through a plane surface. — Focal lines. — Caustics. — Refraction through parallel layers. — Mirage. — Prisms. — Refraction through spherical surfaces. — Lenses. Oblique refraction. — Formation of images by lenses. — Dispersion. — Aberration. — Achromatism. — Rainbows. — Halos. — Theoretical deductions of the laws of refraction. — Hamilton's characteristic function . . 198—227

CHAPTER XVII.

RADIATION AND ABSORPTION: SPECTRUM ANALYSIS. — ANOMALOUS DISPERSION. — FLUORESCENCE.

Equality of emissivity and absorptive power. — Spectrum analysis. — Law of absorption. — Body colour. — Dichroism. — Surface colour. — Metallic reflection. — Anomalous dispersion. — Fluorescence. — Phosphorescence. — Theories of dispersion . . . . . 228—240

CHAPTER XVIII.

INTERFERENCE.—DIFFRACTION.

PAGE

Principle of interference.—Young's experiment.—Fresnel's experiment.—Lloyd's experiment.—Fresnel's biprism.—Coloured interference bands.—Displacement of bands by refracting media.—Interference bands in spectra.—Colours of thin plates.—Newton's rings.—Colours of mixed plates.—Colours of thick plates.—Diffraction.—Effect of a rectilinear wave.—Effect of plane and spherical waves.—Diffraction at a straight edge.—Diffraction at a narrow slit.—Diffraction at a circular aperture.—Zone plates.—Diffraction at an opaque disc.—Coronæ.—Young's Eriometer.—Diffraction gratings . . . 241—264

CHAPTER XIX.

DOUBLE REFRACTION.—POLARISATION.

Double refraction.—Huyghens' construction.—Special sections of the surface.—Polarisation.—Laws of polarisation by reflection and refraction.—Direction of vibration in polarised light.—Reflection and refraction of polarised light.—Plane, circular, and elliptic polarisation.—Nature of common light.—Metallic reflection.—Double refraction in biaxal crystals.—Fresnel's theory of double refraction.—Conical refraction.—Interference of polarised light.—Colours of crystalline plates.—Haidinger's brushes.—Artificial production of the doubly refracting structure.—Rotatory polarisation.—Polarising prisms . . 265—295

CHAPTER XX.

THE NATURE OF HEAT.

Radiant heat.—Its identity with light.—Heat in material bodies.—Hypothesis of molecular vortices . . 296—300

CHAPTER XXI.

RADIATION AND ABSORPTION OF HEAT.

Prévost's theory of exchanges.—Stewart's and Kirchhoff's extension of Prévost's theory.—Laws of radiation of heat at constant temperature.—Heat spectra.—Radiation at different temperatures.—Solar radiation.—Radiation from moving bodies . . . . . 301—309

CHAPTER XXII,.

EFFECTS OF THE ABSORPTION OF HEAT: DILATATION AND ITS PRACTICAL APPLICATIONS.

PAGE

Temperature.—Dilatation of solids.—Dilatation of liquids.—Dilatation of gases.—Absolute zero of temperature.—Measurement of temperature . . . 310—323

CHAPTER XXIII.

EFFECTS OF THE ABSORPTION OF HEAT: CHANGE OF TEMPERATURE AND CHANGE OF STATE.

Unit of heat.—Specific heat.—Thermal capacity.—Specific heat of solids and liquids.—Law of Dulong and Petit.—Specific heat of gases and vapours.—Change of molecular state.—Latent heat.—Fusion and solidification.—Latent heat of fusion.—Evaporation and condensation.—Latent heat of vaporisation.—Formation of dew.—Continuity of the liquid and gaseous states.—Critical temperature.—Solution.—Freezing mixtures.—Dissociation and chemical combination . . . . . 324—343

CHAPTER XXIV.

CONDUCTION AND CONVECTION OF HEAT.

Conduction.—Conductivity.—Measurement of conductivity.—Conduction through the earth's crust.—Conduction in crystalline bodies.—Conduction in liquids and gases.—Convection . . . . . 344—354

CHAPTER XXV.

THERMODYNAMICS: HEAT AND WORK.

Mechanical equivalent of heat.—First law of thermodynamics.—Carnot's complete cycle of operations.—Carnot's reversible cycle.—Reversibility the test of perfection.—Second law of thermodynamics.—Absolute temperature.—The indicator diagram and its applications.—Entropy.—Total, available, and dissipated energy.—Thermodynamic motivity . . . . . 355—369

CHAPTER XXVI.

PAGE
THERMODYNAMICAL RELATIONS . 370—376

CHAPTER XXVII.

ELECTROSTATICS.

Electrification by friction.—Conductors and non-conductors.—Fundamental phenomena presented by electrified bodies.—Positive and negative electricity.—The gold-leaf electroscope.—Electrification by contact and by induction.—Electric quantity.—Continued production of electricity.—The electrophorus.—Law of electric attraction and repulsion.—Electric potential.—Electromotive force.—Capacity.—Condensers.—Specific inductive capacity.—Distribution of electricity on conductors.—Electric density.—Electric images.—Lines of electric force.—Electric induction.—Tubes of induction.—Electric energy.—Electric absorption.—Disruptive discharge.—Atmospheric electricity.—Pyro-electricity.—Electrification by contact.—The electrometer.—Electric machines . . . . . 377—412

CHAPTER XXVIII.

THERMO-ELECTRICITY.

Thermo-electric phenomena.—Laws of thermo-electric circuits.—Variation of the electromotive force with temperature.—The thermo-electric diagram.—The Peltier effect.—The Thomson effect.—Further discussion of the thermo-electric diagram . . . . . 413—423

CHAPTER XXIX.

ELECTRIC CURRENTS.

Convection current between charged conductors.—Flow of electricity in metallic conductors.—Ohm's law.—Kirchhoff's laws.—Electrolytic conduction.—Faraday's laws.—Polarisation.—Ohm's law in electrolytes.—Production of electric currents.—Primary cells.—Secondary cells.—Transformations of electric energy in conducting circuits.—Joule's law.—Measurement of electromotive force, current strength, and resistance . . . . 424—440

CHAPTER XXX.

MAGNETISM.

PAGE

Fundamental phenomena.—North and south magnetism.—Paramagnetic and diamagnetic bodies.—Magnetism a molecular phenomenon.—The law of magnetic attraction and repulsion.—Poles, axis, and magnetic moment of a magnet.—Lines of magnetic force.—Magnetic potential.—Magnetic intensity.—Magnetic induction.—Permeability and susceptibility.—Residual magnetism.—Retentiveness.—Coercive force.—Relation connecting magnetisation and magnetising force.—Hysterēsis.—Effects of vibration and of temperature.—Effects of stress.—Magnetometric measurements.—Terrestrial magnetism.—Theories of magnetism . . . . . 441—462

CHAPTER XXXI.

ELECTROMAGNETISM, ETC.

Oersted's discovery.—Magnetic action of closed electric circuits.—Electrodynamic action on an electric circuit.—Linear circuits.—Circular circuits.—Solenoids.—Ampère's hypothesis regarding magnetism.—Continuous rotation under electromagnetic force.—Electric motors.—Electromagnetic induction.—The dynamo.—Electrokinetic energy.—The galvanometer.—The ballistic method.—Electric and magnetic units . . . . . 463—480

CHAPTER XXXII.

ELECTROMAGNETIC THEORY OF LIGHT.

Magnetic rotation of the plane of polarisation of light.—Hypothesis of molecular vortices.—Hall's effects.—Kerr's effects.—Electromagnetic theory of light.—Electromagnetic waves . . . . . 481—489

CHAPTER XXXIII.

THE ETHER.

Direct action at a distance.—Action through a medium.—Rigidity and density of the ether.—Elastic-solid theory.—Contractile ether.—Electromagnetic ether.—Dilatancy 421—496

INDEX . . . . . 497—501

ERRATA.

Page 17, line 11, *for* 'wrong' *read* 'discordant.'

,, 26, ,, 9, *for* 'the closed curve at' *read* 'a closed curve surrounding.'

,, 28, ,, 8 from foot, *for* 'here' *read* 'there.'

,, 49, ,, 1, *for* 'relative position of P' *read* 'position of P relative.'

,, 149, Equation (2), *for* $\frac{\pi n r^4}{2\mathrm{I}}\theta$ *read* $-\frac{\pi n r^4}{2\mathrm{I}}\theta$.

,, 173, ,, (4), *for* 'ρd' *read* '$d\rho$.'

,, 183, line 12, *for* '63' *read* '64.'

,, 198, ,, 13, *for* 'refraction' *read* 'reflection.'

,, 207, ,, 1, *for* 'μ' *read* '$1/\mu$.'

,, 248, ,, 5, *for* 'Fox,' *read* 'Fox-'.

,, 257, ,, 3, *for* 'e_2, e_2' *read* 'e_1, e_2.'

,, 273, ,, 15, *for* 'XXXIII.' *read* 'XXXII.'

,, 395, ,, 12, *for* $\frac{r'^6}{a}$ *read* $\frac{r'^6}{a^6}$.

A MANUAL OF PHYSICS.

CHAPTER I.

INTRODUCTORY: THE PHYSICAL UNIVERSE.

1. All processes which occur in the universe may be classified as physical or as non-physical. They appertain essentially on the one hand to dead matter, or, on the other hand, to matter which possesses (or is possessed by) life. This statement, of course, is a mere definition of what is meant by the word *physical*, and is not to be regarded as being in any sense the expression of an opinion regarding the nature of life. If, however, in the phenomena associated with living bodies we meet with processes akin to others which occur in the inorganic world, we regard them as being purely physical; while if, in these phenomena, we meet with processes which are totally dissimilar to any with which we are acquainted in the phenomena of dead matter, we leave their further study to the biologist.

2. It is evident, therefore, that the domain of the physical sciences is of immense extent—of such extent that no one, in the course of the longest life, could hope to master all its known details. At one time, indeed (and that not very remote), the scientist might have said after the fashion of Francis Bacon 'I have taken all knowledge for my realm,' but such a claim would be impossible now. All questions regarding the combinations and interactions of the various kinds of matter are purely physical in their nature, but their study is now left to specialists in the department of chemistry. The investigation and prediction of the motions of the heavenly bodies and the determination of their physical constitution are left to the astronomer, while the configuration of the earth is studied by the geographer. So also the sciences of navigation and ship-designing, of engineering, mineralogy, geology (in large

part at least); meteorology, and so on, are purely physical sciences, the study of which is undertaken by specialists, for all of them could never be fully studied by any one man. Thus with increase of knowledge the domain of the physical sciences has been subdivided, and the terms *Physical Science*, *Natural Philosophy*, or *Physics*, have become restricted in meaning, so that they refer merely to the pure scientific groundwork which underlies all the more practical or more highly specialized physical sciences.

3. In commencing the study of physics we have no concern with purely metaphysical questions regarding the objective reality or non-reality of the universe. We simply assume that it has an existence quite independently of the existence of an observing mind, and then proceed to examine the facts and phenomena which make up its entirety. But, before entering upon any detailed examination, it is well that we should take a glance at our subject as a whole in order to learn something of its scope and of the mutual relations of its various parts: and this just for the same reason as that which makes it desirable for the traveller in an unknown country to examine it first from the vantage-ground of some commanding height, so that he may carry with him in his future wanderings therein a clear mental picture of its disposition.

4. A most obvious difficulty meets us at the very outset. How are we to distinguish between that which has true existence and that which has only the appearance of it—between the true landscape, as it were, and the mirage? The mirage seems to the observer to be as real as the reality. What, then, is to be the test of true existence?

In addition to the assumption of the true existence of the physical universe, it is assumed that in this universe *there is nothing which does not occur according to law*. But this assumption is not left without support. The mere possibility of the existence of the so-called exact sciences may be taken as evidence of its truth. Keeping this idea in view, we see that no real thing can appear in, or disappear from, the universe in an arbitrary manner. Therefore we cannot regard anything as a reality unless we can prove it to be constant in amount, that is (to use the ordinary scientific expression), unless we can show that it possesses the property of *conservation*. CONSERVATION is our great test of reality.

This test being applied, it is found that there are two, and only two, classes of things in the physical world—MATTER and ENERGY—which sustain it.

5. Everyone knows what is meant by the term the 'matter' or 'material' of which a body is composed, so, in the meantime, no attempt at a definition is necessary.

At first sight it might appear that matter is certainly *not* conserved. If we weigh a lump of pure limestone, so as to determine the amount of matter in it, and then heat it sufficiently, we find that its weight has become less during the process. It would seem that matter has been lost. But what has actually occurred is the decomposition of the carbonate of lime, by the application of the heat, into lime and carbonic acid gas. The latter, being colourless, passes off unnoticed, and the second weighing gives only the weight of the lime. By proper means the weight of the gas may be determined, and it is then found that the original weight of the limestone is equal to the sum of the weights of its constituents. No matter has been lost in the process. And in *all* chemical processes, however complex, the same result holds: indeed, the science of chemistry is a possibility only in virtue of the strict conservation of matter. Therefore we say that matter is a real thing.

There are many kinds of matter which differ from one another to a greater or less extent in their various physical properties. These properties will be considered in detail in subsequent chapters, and their variations from one to another of a few of the most important or most peculiar substances will be indicated; but the enumeration, classification, and investigation of the various substances in nature belong more to the science of chemistry than to that of physics. One special substance, which pervades all others and extends throughout the whole of the visible universe, possesses such extraordinary properties, and is of such immense importance physically, that it must receive separate treatment. (Chap. XXXIII.)

6. In addition to the property of conservation, matter is characterised by passivity or inertness. It is said to possess INERTIA. In other words, a material body can do nothing of itself. If at rest, it cannot move unless something outside of itself sets it in motion. If moving, it cannot come to rest or alter its motion in any way unless something external to it produce that effect. This property and its consequences will be discussed under Newton's 'First Law of Motion' (Chap. VI.). It is the distinguishing characteristic of matter. One kind of matter can unite with another kind so as to produce a compound, differing entirely, it may be, in its properties from both of its constituents: but, without something external to the matter, no such combination or change could occur.

7. There is every reason to believe that the inertia of matter

cannot be overcome, that is, that the motion of matter cannot be altered, unless motion is imparted to it from some other portion of matter, whether by direct collision or otherwise. It is commonly said, in such a case, that the second body 'does work' upon the first, or that the first has 'work done' upon it by the second. Thus 'the doing of work' involves essentially the production of motion. We might quite consistently assert that the second body possesses 'work,' and that it imparts 'work' to the first—the transference of work being that process which, in ordinary language, is called the performance of work. But it is usual and convenient to adopt the term ENERGY instead; for, although in all likelihood the possession of work or energy necessarily involves motion of the material system which possesses it, the moving system may not be evident to our senses, in which case it is convenient to speak of the energy as existing *potentially* in some connected system which may be at rest at the time, but which can be set in motion by having energy imparted to it from the other. (See next section.)

Of two bodies moving with a common speed, that one with least mass (quantity of matter) can do least work, and the same is true of the slower moving of two bodies which have the same mass. The work tends to vanish either as the mass or the speed becomes indefinitely small; and, so far as experiment (which is the only permissible test) shows, it does not depend upon anything else than the mass and the speed.

Each unit of mass, moving with a given speed, is found to possess the same amount of energy; and, therefore, the energy of any body is proportional to the total quantity of matter which it contains.

Now suppose that two bodies, whose masses are equal in amount, are moving in the same straight line towards each other with equal speeds. Each possesses an equal amount of energy, E (say). We may assume that the result of the impact is that each body is brought to rest, so that the quantity of energy $2E$ has been expended in stopping the forward motion of the two masses. Next, let one of the bodies be at rest while the speed of the other is doubled. The relative speed of the two is unaltered, and hence the energy expended in impact is still $2E$. But experiment shows that the two bodies will now move together with half the speed of the single one, so that each body still possesses energy E. Therefore the single body, moving with double speed, possesses an amount of energy $4E$.

By this experiment, and by other similar experiments conducted under varying conditions regarding the speeds of the moving bodies,

it may be proved that the energy of a moving body varies directly as the square of its speed.

If E be the energy of the moving body, we may express the results just obtained by means of the equation

$$E = kmv^2,$$

where k is a constant. The value of k is purely arbitrary, depending on our choice of the unit of energy, but it will be shown subsequently that it is convenient to choose the units, so that the value of k is $\frac{1}{2}$. Thus

$$E = \tfrac{1}{2}mv^2.$$

8. Energy, when it is regarded as in the preceding paragraph, is called energy of motion or *kinetic energy*. But, as already indicated, we do not always perceive the system to which such energy is communicated. The kinetic energy which the visible system still possesses becomes less and less, and at last, in a certain position of the system, it may entirely vanish. The kinetic energy remaining at any instant may obviously be expressed in terms of the position of the system, and so also may the energy which has been given away from it. Therefore, on the assumption, made in § 7, that the gain of energy of the invisible system is equal to the loss of energy of the visible one, we can represent the kinetic energy of the invisible system in terms of the position of the visible one. This is always done when the two systems are so connected that, when left to themselves, the energy will be recommunicated to the original one, and so we speak of energy of position or *potential energy*.

As a special example we may consider the case of a bullet fired vertically upwards. The further it rises the less its speed becomes, and the less work it can do in overcoming obstacles. At last it comes to rest, and is totally devoid of kinetic energy. But we have only to let it fall down again, and (neglecting the resistance of the air) we find that its energy of motion, when it again reaches the ground, has the same value as at first. Hence, instead of saying that as the ball loses energy some connected system gains it, we, for convenience, say that as it loses kinetic energy it gains potential energy.

We do not yet know what this connected system, in the case of gravitation, is. If Le Sage's hypothesis of ultra-mundane corpuscles (Chap. VIII.) were true, the kinetic energy of a ball, projected upwards against gravity, would be transmuted into kinetic energy of the corpuscles.

We may distinguish a number of forms of energy, all of which

can be classified under the two main types just defined. Kinetic and potential energy of visible portions of matter have been already considered. There is also kinetic energy of invisible portions of matter (Chap. XX.), as in the case of a body which is sensibly hot. And potential energy also exists on a similar scale, as in the case of the so-called Latent Heat (Chap. XXIII.). Other examples appear in the molecular motion of gases, and in the transmission of vibrations through an elastic medium, etc. Again, two oppositely electrified bodies attract each other, and so have potential energy of electrical separation. And, when electricity flows along a conductor, the energy of electricity in motion becomes evident. Also two chemical substances which tend to combine and form a compound substance are said to have potential energy of chemical separation. Lastly, two magnets have potential energy relatively to each other, and work can be indirectly produced by means of the motion of magnets.

9. At this point we are led to regard that characteristic in the possession of which energy differs totally and fundamentally from matter. While matter is essentially passive, energy is constantly in a state of change. It is constantly being handed on from one portion of matter to another, and is ever being changed from one to another of the forms above indicated. It is said to possess the property of TRANSFORMATION.

Only in virtue of this property can we recognise its existence. We could never have known that a moving cannon ball possessed energy had we never seen its destructive effects. We would have been ignorant of the energy of an electrified thunder-cloud if we had not seen the production from it of light, heat, sound, and mechanical effect. *How* energy is passed on from one material system to another, and *how* it changes from one form to another, are questions to which no final answer can at present be given.

A simple example of the transformation of energy is furnished by the motion of an ordinary pendulum. At the lowest part of the swing the energy is entirely kinetic; at the highest part it is entirely potential; and, in intermediate positions, it is partly kinetic, partly potential.

A somewhat more complex example occurs in the transmission of a message by telephone. There is first the energy of vibratory motion of the air when the sound is produced. This vibratory motion is communicated to the metallic diaphragm of the telephone. But the diaphragm is magnetised by induction, and so its motion causes alterations in the intensity of magnetisation of the magnet. These alterations in the magnetisation produce electric currents in the wire coiled round the magnet, and these currents produce similar

alterations of magnetisation in the magnet of the receiving telephone, and so similar motions of its diaphragm ensue. Consequently similar sounds are heard at the receiving instrument.

By the consideration of such special examples we are led to the conclusion that any form of energy may, directly or indirectly, be changed into any other. Many evidences of this will appear in subsequent chapters.

10. During all its changes and transferences one thing is evident regarding the energy in the universe—the total amount of it is unalterable. This is made clear by the fact that strict 'mechanical equivalents' of heat and the various other forms of energy are obtainable (Chap. XXV.). Energy, like matter, possesses the property of CONSERVATION. The swings of a pendulum, which has been set in motion and then left to itself, gradually die away, and finally vanish. But if no energy were lost because of the communication of motion to the air, and if none were lost because of friction at the points of support, or because of vibrations set up in the supporting framework, etc., the motion would go on for ever. That is to say, the energy communicated to other bodies up to any instant, together with the energy still possessed by the pendulum at that instant, is equal in amount to the original quantity. And the same is true of any other system. Therefore, since it is conserved, we must regard energy as having real existence.

11. A question of the deepest importance to mankind arises in connection with the transformation of energy. Are all forms of energy equally transformable? When energy is changed from one form to another, can it with equal readiness be changed back again into the original form? If not, it necessarily follows that the whole amount of energy in the universe will gradually assume that particular form which is least transformable. Observation and experiment have shown that there is one form into which all others are gradually and permanently changing; and that form is the energy of molecular motion known as *heat*. But there is a constant tendency towards diffusion of heat, so as to produce uniformity of temperature; and when uniformity of temperature is arrived at, no mechanical work can be produced from the heat. The total amount of energy in the universe will be the same as before, in accordance with the principle of conservation; but none of it will be available for the production of mechanical work. This principle of the loss of availability of energy is technically known as the principle of DISSIPATION, or (preferably perhaps) DEGRADATION of energy.

Examples of degradation of energy occur everywhere in nature. No stone falls from a cliff, no storm arises or ceases, no flash of

lightning or peal of thunder occurs, no wave breaks upon the shore, without a diminution of the possible amount of useful work obtainable for man by natural processes.

In accordance with this principle, potential energy of visible portions of matter tends towards a minimum value, *i.e.*, tends as far as possible to take the form of kinetic energy. But further consideration of this subject must be deferred in the meantime.

12. Nothing which is not either matter or energy is conserved—at least, in the same sense as that in which we assert conservation of these things. Matter and energy are both signless quantities. We might assert conservation of a quantity which may be positive or negative, provided that, when a new positive amount of it is produced, an equal negative amount necessarily appears. In this case the total algebraic sum of the quantity is constant. But if we regard either the positive portion of it or the negative portion of it, we find that the amount of either portion may be perfectly arbitrary; and this is not the sense in which conservation is asserted of matter and energy.

In the new sense alluded to, we speak of the conservation of momentum (Chap. VI.), and sometimes even of the conservation of electricity.

CHAPTER II.

THE METHODS OF PHYSICAL SCIENCE.

13. The whole body of scientific knowledge has been obtained by one or other of two methods—*observation* or *experiment:* nor can strict knowledge be obtained in any other way. The first scientific investigations ever made must have been of a purely observational type, and were very probably astronomical in their nature. In making observations, we notice the positions of objects and the sequence of events; and we attempt, then, to make out relations among them. In this way arose the still-extant grouping of the stars into various constellations, and — greatest perhaps of all examples—the discovery by Kepler of the laws which regulate the motions of the planets. But, when we alter at will the conditions attending certain phenomena, so as to discover the consequent alterations produced in the phenomena, we are said to experiment. It is true, indeed, that we cannot always draw a hard and fast distinction between observation and experiment. Thus, in calculating the speed of light from observations upon the satellites of Jupiter, although we do not ourselves alter any conditions, yet we purposely take advantage of alterations which occur naturally.

By such means we first of all obtain mere series of facts often without any mutual connection whatsoever, and, not infrequently, so grouped as to suggest false relations. The next duty of the scientist is to group these isolated data after a definite system, to co-ordinate the facts with the object of subsequently discovering the true relations which subsist among them; and the greater the power of the observer to detect real resemblances and essential differences the sooner will his ulterior object be attained.

The question of cause and effect next arises. Of two phenomena which appear successively, and no one of which appears without the other, that one which is first evident is usually called the cause of the other, which is said to be its effect. But, obviously, great care

must be taken to avoid any error in such an assertion, for there may be many sources of mistake. In the first place, it is conceivable that two phenomena might appear in invariable succession the one to the other, and yet the true explanation might be that they had a common cause, and were not otherwise connected. The flash of forked lightning and the sound of thunder occur successively; but the sound is due to the explosive expansion of the air heated by the passage of the electricity, while a portion of the light is also caused by this explosive expansion, which compresses the adjacent layers so suddenly as to render them luminous by the excessive heat so developed. Again, the occurrence of one event is frequently necessary, in order that we may perceive another between which and the former there is no connection whatsoever; and, frequently, the effect becomes evident before the cause is noticed. Still further, we observe that events sometimes occur simultaneously. Thus, tornadoes are often due to the sudden heating of large portions of the atmosphere by means of the latent heat given out on rapid condensation of vapour. But the condensation of vapour and the evolution of latent heat occur of necessity at one and the same instant, so that we might with equal propriety refer the tornado to either event as a cause.

In nature there is an apparently endless series of causes. Each event is the cause of another, and was itself produced as the consequence of a preceding event. When a large mass of cloud intercepts the rays of the sun from the underlying atmosphere, the air grows colder, and as it grows colder it contracts. This causes an inrush of air from surrounding regions, which well-known result is expressed by the popular phrase that the cloud or the rain 'draws' the wind. The effects of this motion might be traced out endlessly if our senses were sufficiently acute and our powers sufficiently universal; and so also the various motions preceding the motion of the cloud might be traced.

14. Among his other duties, the physicist has to undertake the investigation of the effects which result from physical conditions. Such an investigation is comparatively simple. He has only to make certain that the effects which he observes are not due to any unnoticed conditions. The converse problem—the investigation of causes—is not by any means so simple. The investigator must first determine the various physical conditions which actually obtain, and he must then find out which of these, if any, are essential to the production of the phenomenon. If three conditions are observed, an experiment must be made in which all of them are present, in order to make it certain that the result really follows. Then three

experiments must be made with the three pairs of conditions. Three more must then be performed with each condition present alone. Lastly, it may be necessary to make another experiment in the absence of all the conditions. In all, eight experiments may be necessary when there are three conditions. If four conditions are present, one experiment with all the conditions present, four with three conditions, six with two, four with one, and one with none, may be required—in all sixteen experiments. With one condition only, not more than two experiments are necessary, and the number is doubled for every additional condition introduced. If only ten conditions existed, more than a thousand experiments would be necessary to completely exhaust all the possibilities. Obviously, science could make little progress were such immense labour a necessity. Fortunately it is not. Past experience and natural instinct indicate to the experimenter the direction in which truth lies, and thus he is often enabled to take a short road to the end in view.

15. One great means by which labour is reduced is the employment of a suitable and probable *hypothesis*. Certain facts are known, and a hypothesis is framed regarding their explanation.

The greater the number of phenomena which a given hypothesis can explain, the greater is the likelihood of its truth. When only some facts fall in with the assumptions while others do not, some modification of the hypothesis must be made. But when new modifications have to be made for every new requirement, it is time to abandon the hypothesis and seek for another and more probable one. A good hypothesis must explain all the facts for the elucidation of which it was framed. It should also explain other known facts, and facts which subsequently become known. But, in such a case, it is customary to speak of it as a *theory*. Above all, a good theory should lead to the prediction of previously unknown facts.

It sometimes happens that different theories are each sufficient for the explanation of known facts. One theory may explain certain phenomena more easily than another can, while in the explanation of others it is more laboured. The logical consequences of the two theories must then be worked out as far as possible, and it will usually be found that at one or more points each leads to an opposite conclusion. Here experiment must step in to determine which conclusion is correct, and so to decide between the two theories. This experimental investigation is termed a *crucial test*. Very prominent examples occur in the theories of heat and light.

The tendency of scientific investigation in the present day is towards the formation of dynamical explanations of all phenomena—

towards the production of theories in which all purely physical phenomena are explained in terms of matter, and the energy which is associated with it.

Mathematical theories form an important class in which the mathematical consequences of the fundamental assumptions are rigidly worked out. When the postulates are merely expressions of known facts, the consequences of such theories may be regarded as strictly true; but in making such a statement, we must remember that all our knowledge is only approximate, being limited by the imperfections of our senses and our instruments, so that the above expression, 'strictly true,' means merely that we cannot detect deviations from the truth. The theory of gravitation is of this kind, and it furnishes us with one of the finest examples of prediction. From irregularities in the motion of the planet Uranus, Adams and Leverrier were led to foretell the existence and indicate the position of the previously unknown planet Neptune.

In other mathematical theories there is merely a partial experimental basis; for example, the dynamical theory of heat or the undulatory theory of light. Again, it is possible to work out mathematical theories of phenomena in which we know that something moves; but we may not know *what* is moving or *how* the motion is propagated. The theories of heat-conduction and of electro-dynamics are prominent examples.

As knowledge advances theory must cease. Some theories will be shown to be false, while the truth of others will be confirmed, in which case they of necessity vanish as theories.

A very important scientific method, which is in essence hypothetical, is known as the *argument from analogy*. When we perceive resemblances between different physical systems or processes, we say that they are analogous; and when any new fact is discovered regarding one of the systems, we are led by analogy to look for something similar in the others. The principle is of extreme importance in experimental work, as it indicates a promising direction for research, and so prevents aimless and often fruitless labour; and, further, the failure of an analogy may be as instructive as its success. There are many analogies between the phenomena of sound and of light; but there is nothing in sound which corresponds to polarisation in light, and the distinction is of fundamental importance.

Another extremely important aid to research is derived from the condition for stable equilibrium. This condition may be expressed as follows: A system is in stable equilibrium, under given physical conditions, when any small variation of one or more of these pro-

duces other variations which would themselves, as causes, produce changes opposite to the first.

[It is easy to see that this statement does express the condition for stable equilibrium. For, if one variation produced another variation which caused further variation of the first kind, this additional variation would cause more variation of the second kind, and so on reciprocally. Therefore the variation, once started, would constantly increase, or, in other words, the presumed state of equilibrium is unstable.]

The equilibrium of a body supported by the hand affords a ready illustration. Increased pressure of the hand upon the body causes an upward motion of the body. Conversely, the independent communication of upward motion to the body diminishes the normal pressure between the hand and it.

Another example is furnished by water, which is physically stable (under ordinary circumstances) below its maximum-density point, and which, at temperatures below its maximum-density point, contracts when the temperature is raised. Hence, in accordance with the above principle, we can assert that sudden diminution of volume caused by the application of pressure will produce a fall in temperature. Again, sudden elongation heats indiarubber; therefore the heating of the stretched indiarubber makes it shrink. Sir W. Thomson has proved experimentally that both these results are true. We shall see subsequently that they follow as consequences of the dynamical theory of heat.

16. In all observations, alike of natural processes and of experimental results, errors of observation are almost certain to arise. Such inaccuracies are as likely to be in excess as in defect, and are much more likely to be small than to be large, while a very large error will practically never occur at all unless the method of observation is an extremely objectionable one. To get rid of these errors we must make a sufficient number of independent observations. In any one observation we do not expect the result to be correct; but there is a certain numerical quantity, called the *probable error*, such that the actual error is as likely to be greater than it as to be smaller than it. If each observation is made under conditions precisely similar to those of another, the probable error of each is the same. In this case we simply take the arithmetical mean of all the observations, and this gives the result which is most likely to be near the true value. But if each observation is not made under precisely similar circumstances, the probable error of each will in general be different, and its value will be known for each from the known experimental conditions. The most probable

value is now found by the method of least squares. As a simple example, let us take the case of three independent observations of a quantity x, which gives the results $x=a$, $x=b$, $x=c$: and let the probable error of b and c be $\frac{1}{n}$th and $\frac{1}{m}$th of that of a respectively. If we now multiply the second and third equations by n and m respectively, we get $x=a$, $nx=nb$, $mx=mc$, where the probable error of the right-hand member of each equation is the same. Multiply again by n and m as before, and we get $x=a$, $n^2x=n^2b$, $m^2x=m^2c$. These equations give

$$x = \frac{a + n^2b + m^2c}{1 + n^2 + m^2},$$

a value which makes the sum of the squares of the errors of the original equations a minimum. If the equations contain more than one quantity subject to error, the same method applies, for the number of observations will usually be much greater than the number of unknown quantities.

In addition to errors of observation there may be errors which tend always in one direction, so that the result obtained is either too large or too small. Such errors are due usually to the instrument or to the method of observation used, and are generally termed *instrumental errors*. Under this heading may be included errors due to peculiarities of the observer, and the correction to be applied is termed the *personal equation*. When only comparative values of a quantity under different circumstances are required, such errors frequently affect each observation alike, and so may be neglected. But, in general, they must be eliminated by varying the instrument and the observer and combining the results as above.

17. Most frequently in physical inquiries we have to investigate the variations of some quantity consequent upon the variation of another. The experiment may often be so arranged as to give a continuous record of the mutual variation of the two, as in the case of the self-registering thermometer and similar instruments; or even a simultaneous and continuous record, as in the case of the rise of water in the wedge-shaped space between two vertical glass plates (§ 122). But, more generally, the results of a few separate experiments are given, each of which records one definite value of the one quantity corresponding to one definite value of the other. From these detached results the law connecting the variations of the two, or, rather, an approximate law must be found, the approximation to be so exact that the result given by the law for intermediate values of the quantities shall not differ from that which may be determined afterwards by experiment by an amount greater than

the possible error of observation. Such a relation is termed an *empirical law*, and the formula expressing it is called an empirical formula.

The formula

$$y = a + b\,(x - x_0) + c\,(x - x_0)^2 + \ldots,$$

is frequently adequate for the close representation of many experimental results. The constant a is the observed value of y when $x = x_0$; and the values of the constants b, c, etc., are found from a series of particular equations obtained from the above by giving x and y simultaneously-observed numerical values. Frequently no more than three terms are necessary. For example, the values of x and y which are contained in the table below are accurately represented by the formula

$$y = 1 + 2\,(x - 1) + (x - 1)^2,$$

or by

$$y = 9 + 6\,(x - 3) + (x - 3)^2, \text{ etc.}$$

If we have obtained by experiment n values of y corresponding to n values of x, we may use as an empirical formula the equation (given by Laplace):

$$y = \left(\frac{y_1}{x - x_1} \frac{1}{(x_1 - x_2)(x_1 - x_3) \ldots (x_1 - x_n)} + \frac{y_2}{x - x_2} \frac{1}{(x_2 - x_1)(x_2 - x_3) \ldots (x_2 - x_n)} + \cdots \right) (x - x_1) \ldots (x - x_n).$$

This obviously gives $y = y_1$, when $x = x_1$, etc., so that all the observed values are accounted for.

When values of y corresponding to equi-different values of x are observed, the symbolical equation

$$\overline{m+n} = \bar{m} + n\Delta\bar{m} + \frac{n(n-1)}{2}\Delta^2\bar{m} + \ldots.$$

is specially useful. Here $\bar{m}$ means 'the numerical value of y, which stands m^{th} in the observed series'; $\Delta\bar{m} = \overline{m+1} - \bar{m}$; and $\Delta^2\bar{m} = \Delta\overline{m+1} - \Delta\bar{m}$. $\Delta\bar{m}$, etc., are called the 'first differences,' and $\Delta^2\bar{m}$, etc., are called the 'second differences' of the observed values of y. For example, let the values of x be the natural numbers, while the values of y are the squares of these. From the tabulated results:

x	1	2	3	4	5	6	7
y	1	4	9	16	25	36	49
Δy		3	5	7	9	11	13
$\Delta^2 y$			2	2	2	2	2

we get, by the formula, for the sixth observed value of y the quantity $\bar{6}=\overline{\bar{3}+\bar{3}}=\bar{3}+3\triangle\bar{3}+3\triangle^2\bar{3}=9+3\times7+3\times2=36$; or $\bar{6}=\overline{\bar{4}+2}=\bar{4}+2\triangle\bar{4}+\triangle^2\bar{4}=16+2\times9+2=36$; and so on.

When the observed values are sufficiently close together, such formulæ enable us to find intermediate values with considerable accuracy, and also to find values altogether outside the experimental range for a short distance; but, if pushed too far, the formulæ will give values differing more and more from the truth.

Instead of seeking for an empirical formula, we might plot a

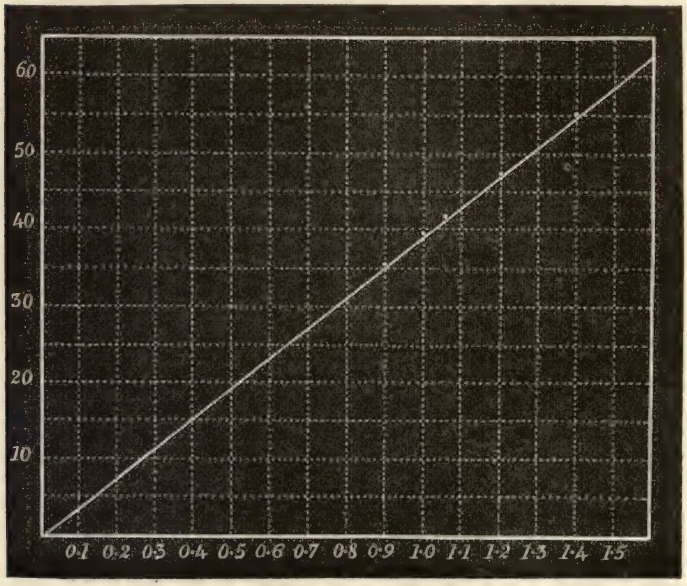


FIG. 1.

curve, the abscissæ of which represent the values of the quantity to which we give arbitrary values, while the ordinates represent the values of the quantity whose variation we observe. The points so obtained will not generally lie on a smooth curve because of observational errors, but a curve drawn freely through them, leaving on the whole as many points on one side of it as there are on the other, will be fairly free from such inaccuracies, and will generally give a better approximation to the truth than the actual points themselves give. This is known as the *graphical method*, and is largely used

by experimentalists. For example, the square of the time of oscillation of a simple pendulum is proportional to the length of the pendulum. Hence, if we make the abscissæ represent lengths while the ordinates represent the squares of the periods of oscillation, we should get a straight line passing through the origin. In Fig. 1, which is drawn from the results of actual experiment, it will be seen that the points, while they all agree very well with each other, do not give the proper ratio of the quantities. The straight line is not inclined at the proper angle. From the close agreement of the different results, we infer that there is little error of observation, but the wrong inclination of the line shows that there must be an instrumental source of error. If a series of experiments were made at parts of the earth's surface where gravity had sensibly different values, we should get a series of straight lines all passing through the origin, but all inclined at different angles. All such series of curves may be regarded as contours of a surface, and this subject is of such importance in physics as to merit discussion in a separate chapter.

Very often the form of the curve obtained by the graphical method indicates at once a suitable empirical formula.

CHAPTER III.

THE THEORY OF CONTOURS, AND ITS PHYSICAL APPLICATIONS.

18. The nature of any quantity is completely known when it is understood *what* units are involved in its measurement, and *how* they are involved. Thus a speed involves the unit of length directly, and the unit of time inversely; an acceleration involves a length directly, and the square of a time inversely. But, when dealing with extension, we have only to consider the unit of length. We say that the extension under consideration has one, two, or three, etc., *dimensions*, according as the unit of length is involved to the first, second, or third, etc., power. A line has only one dimension, for only one number, with the proper sign attached, is required to completely specify the relative position of two points on the line. A surface has two dimensions, for two directed lengths define the position of a point on it with reference to any other point taken as origin. Thus we speak of one point on the surface of the earth as being so much north or south, and so much east or west of another. Three directed lengths determine the relative position of two points in space, that is, in extension of the third order. Thus we speak of the length, breadth, and thickness of a solid.

It must be observed that three *lengths* are not *necessary*. *One* of the given conditions must be a length, but the others might be angles. For example, instead of saying that one mountain-top is so far north or south, so far east or west, and so much higher or lower, than another, we might give the distance between the two peaks and their relative altitude and azimuth.

The intersection of any surface, which has a constant characteristic, with the surface of a solid is called a *contour-line*. A good illustration is furnished by the contour-lines in an Ordnance map. Any such line is the intersection of a surface, all points of which are at a constant height above sea-level, with the surface of the solid earth. An Ordnance map gives a very good idea of the distribution of places on the earth's surface as regards height, as well

as with reference to latitude and longitude, and the information is more and more minute as the number of lines is greater. In other words, contours enable us to represent on a plane surface the mutual relations of *three* quantities.

This gives the key to the great importance of the theory of contours in physical science. For, if the physical condition of a substance is completely defined when the simultaneous values of three of its properties are given, we can construct a solid the surface of which represents all possible conditions of the substance—just as we can construct a model of the earth's surface in terms of latitude, longitude, and height above sea-level.

19. We may extend the above conception, and state, generally, that the *contour of an object of* n *dimensions, existing in extension of the* $(n+1)^{th}$ *order, is its intersection with an object of* n *dimensions at every point of which some quantity has a constant value.* It is, therefore, of $(n-1)$ dimensions.

Since, in extension of the $(n+1)^{th}$ order, we may have objects of less dimensions than the n^{th}, we might make the further development that in extension of the $(n+1)^{th}$ order we may have contours of all positive dimensions up to the $(n-1)^{th}$ inclusive. Indeed, there is no reason why we should not consider contours of n dimensions in space of $(n+1)$ dimensions. Hence, in ordinary extension, we may have point, curve, and surface contours. The contours of a curve are points; of a surface, curves; of a solid, surfaces; of a four-dimensional object, solids; and so on.

The properties of four-dimensional extension, or even of extension of the n^{th} order, can be treated mathematically; but, from want of experience, it is impossible to imagine the nature of such extension.

20. By means of contour-points the nature of curves may be exhibited in diagrams consisting of straight lines only. For we may intersect a given curve by curves, along each of which some quantity has a constant value, and then project the points of intersection upon any straight line.

Consider first a plane curve, and, for convenience, let its plane be taken as that in which two co-ordinate quantities, x and y, are measured in perpendicular directions from the same origin. Let $f_n(x, y)=0$ be the equation of the given curve, where the suffix denotes the degree of the equation. The equations to the curves along which some quantity, say c, is constant, may be written in the general form, $\phi_n(x, y, c)=0$. In the different curves of the system c has different values. As a particular example, the curves might be circles of different radii. Again, the equation might be

$\phi_1(y, c) = o$, which represents a series of lines parallel to the axis of x. This is the simplest case which we can consider, and, at the same time, the most useful. The curves in Figs. 2 and 3 are intersected by lines parallel to the axis of x, and the points of intersection are projected upon the axis, and are designated by numbers which give the various values of y corresponding to the given values of x. If the curve be continuous, a *maximum* or *minimum* value of y exists between two equal values. It is a maximum if y first increases and then diminishes as x increases continuously from its least value corresponding to the given value of y. It is a minimum if y first diminishes and then increases. The *steepness of slope* is shown by the closeness of the contours for equal increments of y, and its *direction* is shown by the order in which the values of y occur as regards numerical magnitude when x increases.

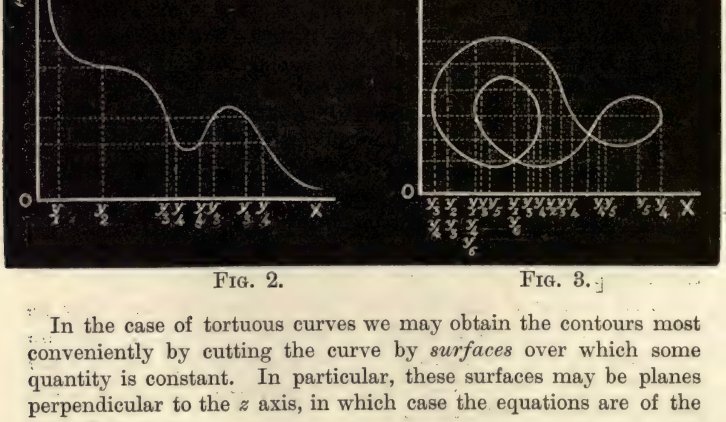


FIG. 2. FIG. 3.

In the case of tortuous curves we may obtain the contours most conveniently by cutting the curve by *surfaces* over which some quantity is constant. In particular, these surfaces may be planes perpendicular to the z axis, in which case the equations are of the form $f_1(z, c) = o$.

The position of a moving point in space is obviously representable by a tortuous curve. Its position at any time can be got from the curve if the value of the time in terms of one of the co-ordinates, say z, is known; for we should then only have to cut the curve by the plane ${}_1f_n(z, t) = o$, where t represents the time, and ${}_1f_n$ is a functional symbol, showing that the equation is of the first degree in z, but may be of any degree in t. [This condition is rendered necessary by the fact that the point must be in one definite

position at a given time, but may occupy the same position at different times.] If a number of such curves are simultaneously traced out in space by moving material points, we can obtain the *diagram of configuration* of the material system at any time by cutting the curves by planes corresponding to that time, and projecting the points of intersection upon the parallel co-ordinate plane.

It is evident that, in general, the plane which corresponds to a definite time will be different for each curve. The disadvantage so entailed may be got rid of by the employment of trilinear co-ordinates to indicate the position of the point when the time is given. If the curve be cut by any plane, the distances of the point of intersection from three intersecting straight lines in that plane give the x, y, z co-ordinates at the corresponding instant. The value of the time might be given by the distance of the plane from a fixed plane parallel to it.

The curves which are, on this system, taken to represent the

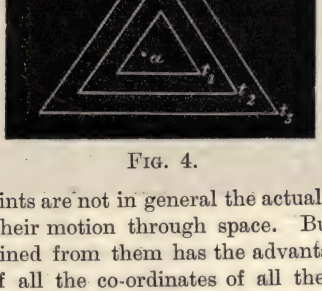


FIG. 4.

positions of the points are not in general the actual curves traced out by the points in their motion through space. But the diagram of configuration obtained from them has the advantage of showing at once the values of all the co-ordinates of all the points, whereas, in the Cartesian system, this could not be done without projection on all the co-ordinate planes. The triangles of reference, though they may be similar, are not generally of the same magnitude. A difference in magnitude is necessary, in order to represent varying values of the co-ordinates. In Fig. 4 the triangles are similar, and equidistant, while the point a is fixed; hence the diagram represents the linear motion of a point. In general there must be a different set of triangles for each separate point whose motion is to be indicated.

By the aid of such a diagram, the *diagram of total displacements* (§ 40) in a given time may be constructed. And by taking the displacements in one n^{th} part of the unit of time (n being

indefinitely large), and magnifying them n times, the *diagram of velocities* can be got. Similarly, the diagrams of *accelerations*, *forces*, and so on, may be represented as the contours of curves. The curves, from which the diagrams of velocities, etc., are obtained, are, it is almost needless to remark, different from the original curves which represent the positions of the moving points. In the case of velocities, they are the hodographs (§ 48) of the original curves on this trilinear system of reference.

As another example of the use of tortuous curves, we may consider two quantities, x and y, connected by the equation $y^2 = ax$, which gives $ydy/dx = a/2$ (§ 30). We may now take dy/dx as a third co-ordinate quantity, and so obtain a tortuous curve. For example, if, in this case, y represents the time during which a body

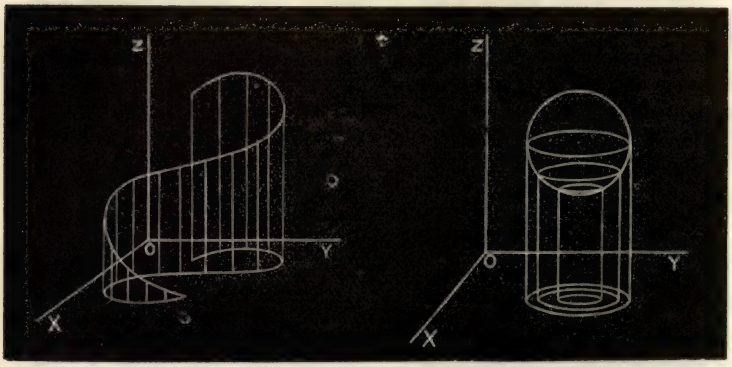


FIG. 5. FIG. 6.

has been falling from rest under gravity, and if x represents the space described from rest, the velocity acquired is represented by the reciprocal of the third co-ordinate quantity.

21. If any curve be cut by planes parallel to that of (x, y), and if the various points of intersection be projected on any one of these planes, say $z = o$, the contour-points so obtained will evidently lie on a definite line, and the line will be more accurately indicated in proportion as the number of intersecting planes is greater and their mutual distance is less. It will be given without any break of continuity by projecting every point of the curve upon the plane $z = o$. But such a line may be regarded as the intersection, by the plane $z = o$, of a cylindrical surface whose generating lines are parallel to the z-axis and are drawn from the given curve to meet that plane. Now this satisfies our definition of a contour-line, for it is the

intersection of a given surface by a surface over which z is constant (zero). A cylindrical surface supplies the simplest diagram of contour-lines. The contours are all superposed in the diagram, but are not in general conterminous. The only case in which they would be conterminous is (Fig. 5) that in which the same values of the x and y co-ordinates of a point on a curve correspond to different values of z.

In the case of a non-cylindrical surface, no part of the contours will be superposable in general. The contours of a hemisphere, for example, are concentric circles (Fig. 6). And, just as, in the case of contour-points, the steepness of slope of the curve is indicated by the closeness of the contour-points on the x-axis for equal increments of y, so, in the case of contour-lines, the steepness of slope of the surface is indicated by the closeness of the contour-lines for

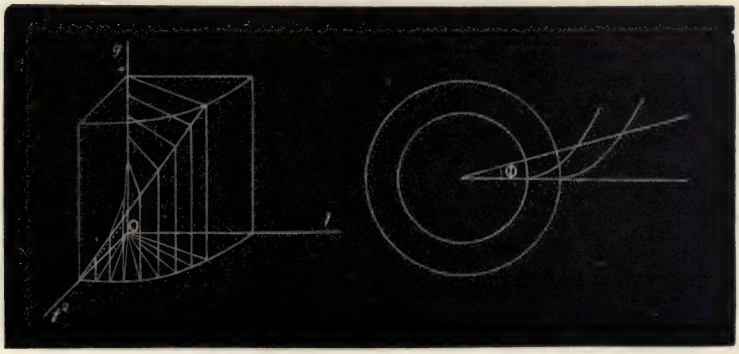


FIG. 7. FIG. 8.

equal increments of z. The contours are closer when their radii are large.

The contours of a right circular cone are also concentric circles, but they are at equal distances apart for equal increments of z.

22. As an example of a surface, the contours of which may be used to indicate certain physical properties, we may consider that one whose equation is

$$z = 4\pi^2\, y/x.$$

If y represents the length of a simple pendulum, while x represents the square of its time of oscillation, we know that z represents the value of the acceleration due to gravity. The surface (Fig. 7) may obviously be supposed to be produced by the motion of a straight line which intersects the z-axis and is always perpendicular to it,

and which rotates uniformly about that axis while it moves at a constant rate along it. The contours (by planes perpendicular to the axis of z) are straight lines passing through the origin and variously inclined to the x-axis (§ 17).

Again, the intrinsic equation of the circle is

$$s = a\phi,$$

where a is the radius and ϕ is the angle between the radius vector and the initial line. Hence the intrinsic equation of one involute is

$$s' = \frac{a}{2}\phi^2.$$

This involute is the one which meets the circle at the position from which ϕ is reckoned, and s' is measured along it from this point. If we consider a to be the mass of a moving body, and ϕ to be its speed, s and s' are respectively its momentum and kinetic energy. In Fig. 8 two circles, with their involutes satisfying the above condition, are drawn. The curves may be regarded as the contours of a right circular cone, together with an associated surface which is formed so that its intersection by any plane parallel to that of the diagram is an involute of the circle in which the cone is cut by the same plane.

The acceleration and speed of a body falling under the action of gravity, and the space passed over by it, are given by the known equations (§ 42)

$$a = g$$
$$v = V + gt$$
$$s = c + Vt + \tfrac{1}{2}gt^2.$$

Hence these various quantities can be represented also by the contours of the surfaces just considered, t taking the place of ϕ, and g taking that of a in the former equations. The new terms which appear present no difficulty.

Again, in a thermo-electric circuit (Chap. XXVIII.), composed of two dissimilar metals, the electromotive force E is given, in terms of t the difference of temperature of the two junctions, by means of the formula

$$E = a + bt + ct^2.$$

Also the thermo-electric power e is given by the equation

$$e = b + 2ct.$$

Hence these quantities are also representable by means of the same surfaces.

23. The contour lines which are most familiar to us are those

formed by the intersection of level surfaces with the surface of the earth. The line of sea-board is one such contour line. The numbers marked upon maps or charts which give the height above, or depth below, sea-level indicate contour-points. When the points are taken sufficiently close together, and continuous curves are drawn through points of constant height, we get contours, as in the Ordnance Survey maps. Such a contour coincides very closely with the contour-lines formed by level surfaces. They do not exactly coincide because of the non-spherical shape of the earth, and because of its rotation, etc. But the assumption that contour-lines of constant *level* are lines of constant height over sea-level will not introduce appreciable error, so long as the area on which they are drawn is

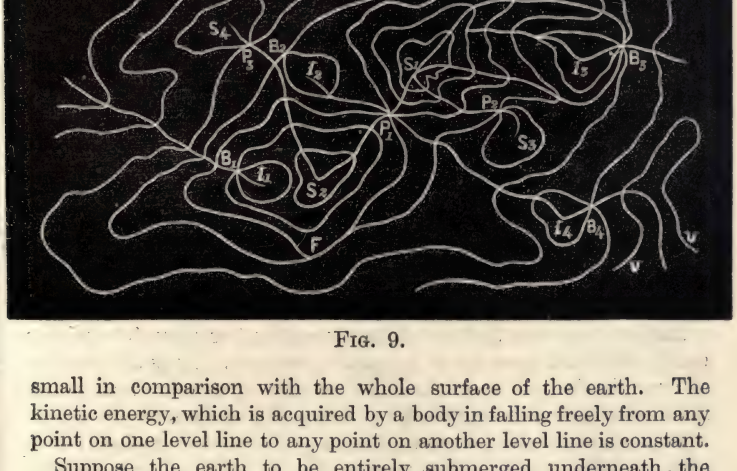


FIG. 9.

small in comparison with the whole surface of the earth. The kinetic energy, which is acquired by a body in falling freely from any point on one level line to any point on another level line is constant.

Suppose the earth to be entirely submerged underneath the surface of water, so that we have only one region, and that a region of depression below the surface of the water, to consider. If we suppose further that the water is slowly absorbed by the solid matter of the earth, regions of elevation will be formed gradually, until finally we shall have again only one region, and that a region of elevation. Before a region of elevation is formed, we have a *summit* appearing above the water-level; and when the water subsides out of a region of depression, we have a lowest-point or *imit* appearing.

The number of regions of elevation and depression may vary in two ways. Two regions of elevation may run into each other as the water sinks. The point where they first meet is termed a *pass* or *col* (see Fig. 9; P_1, P_2, etc.). Again, a region of elevation may throw out arms which run into each other, and so cut off a region of depression. The point where they first meet is termed a *bar* (B_1, B_2, etc.). The contour-line for a level immediately underneath that corresponding to the bar has a closed branch within the region of depression cut off. Thus the closed curve at I_4 is part of the contour-line UV. In the map of such a country, a pass occurs at the node of a figure-of-eight curve (or *out-loop* curve, as Professor Cayley has termed it), while a bar occurs at the node of an *in-loop* curve. If, in the diagram, the Ps represented bars and the Bs represented passes, the map would be that of an inland basin; so that, in the map of such a country, a pass is represented by the node of an *in-loop* curve, and a bar corresponds to the node of an *out-loop* curve. If there were any advantage in having passes and bars always indicated by the node of the same kind of curve respectively, this could be attained by affixing the positive sign—not constantly to the region on the same side of the level surface but—to the region towards which (or from which) the surface is moving at any instant.

As a particular case, two regions of elevation may run into each other at a number of points simultaneously. Of these points, one must be taken as a pass and the others as bars. Singular points may also occur, when, for example, three or more regions of elevation meet. Such points are called *double*, *treble*, etc., passes. Multiple bars may similarly occur.

Before a pass can be formed there must be two summits, and for every additional pass there is another summit. Thus the number of summits is one more than the number of passes. So also the number of imits is one more than the number of bars.

Slope-lines are lines drawn at right angles to the contour-lines; and, evidently, the steepness of a district is indicated on a map by the closeness of the contours. Two kinds of slope-lines are of special importance. These are the slope-lines drawn from summits to passes or bars, and from passes or bars to imits. The former can never reach an imit, and are termed *water-sheds*. The latter can never reach a summit, and are called *water-courses*.

A perpendicular precipice is indicated on a chart by the running together of two or more adjacent contour-lines (F). An overhanging precipice is indicated by the lapping of the upper-level line over a lower-level line.

24. Since we can represent the physical state of a substance with regard to three quantities by means of a surface, it follows that we can deduce from the contours of the surface the nature of the variation of the properties of the substance—the methods being identical with those of the preceding paragraph. Let us take, as a particular example, the thermo-dynamic surface which represents the state of water-substance with regard to volume, pressure, temperature, entropy, and energy (Chap. XXV.). If we construct the surface in terms of any three of these quantities, the value of the remaining two at any point of the surface may be given by contour-lines. The model of the surface, having volume, entropy, and energy measured along the axes, has been constructed by Clerk Maxwell, and is explained and figured in his 'Theory of Heat.' We shall

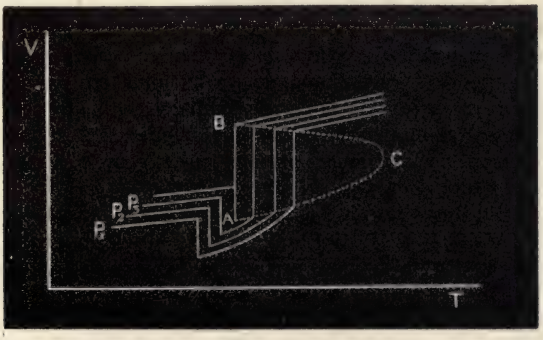


FIG. 10.

consider the surface representing directly volume, temperature, and pressure. This surface was first studied by Professor James Thomson.

Suppose the surface to be cut by a plane of constant pressure, say P_1. We thus get a contour-line, the general nature of which is indicated in Fig. 10. At a low temperature the volume is small, the substance being in the solid state. As the temperature rises, the substance expands until liquefaction occurs. The volume then diminishes without rise of temperature, until the substance is completely liquefied. The temperature then rises while the volume diminishes, until the maximum-density point is reached. After this, expansion accompanies rise of temperature up to the boiling-point. At this stage the volume rapidly increases, while the temperature remains steady until the substance is entirely in the gaseous state. Beyond this point both increase together. The contours for slightly less pressures (P_2, P_3) are approximately parallel to P_1, but lie entirely

above it; the reason being that at a given temperature the volume increases as the pressure diminishes, while the freezing-point is lowered, and the boiling-point is raised, by pressure. The freezing and boiling points approach, and finally coincide as the pressure diminishes. At lower pressures the substance changes directly from the solid into the gaseous state. The line AB in the figure above indicates the *triple-point* temperature, that is, the temperature at which portions of the substance in the three states—solid, liquid, and gaseous—can exist together in equilibrium. The increase of volume on vaporisation continually diminishes as the pressure is raised, until finally (at C) the process of vaporisation ceases. The temperature at which this occurs is called the *critical temperature*. There may also be a critical temperature for the solid-liquid condition. That is to say, there may be a temperature below which no amount of pressure will lower the freezing-point sufficiently to admit of liquefaction.

The contour-lines which are obtained by cutting the surface by planes of constant temperature are called *isothermals*. If the temperature be above the triple-point, but below the critical-point, while the substance is in the gaseous condition, increase of pressure is accompanied by decrease of volume, until the liquefaction commences. At this stage the volume decreases without variation of pressure, until all the substance is liquefied. After this, a very great increase of pressure is required to produce a very small decrease of volume. Two such isothermals (*not* those of water-substance, see § 278) are represented in Fig. 11. The form of an isothermal below the triple-point shows that the solid state is intermediate between the gaseous and the liquid states. As the pressure increases the volume decreases, until the point of sublimation is reached. The pressure then remains constant while the volume diminishes, until all the substance is solidified. Then the volume decreases slowly on increase of pressure, until liquefaction commences. Here the pressure becomes constant, while the volume diminishes until all the ice is melted; after which the volume again decreases slowly on rise of pressure. Thus, here are two kinds of isothermals having their transition-stage at the triple-point temperature. As already indicated, the triple-point pressure occurs at the transition between two kinds of lines of equal pressure, for the liquid condition ceases to be possible at higher pressures. The form of the isothermals beyond the critical temperature is indicated in Fig. 11. It is quite possible that, as Professor James Thomson has suggested, the true form of the isothermals below the critical temperature does not include a

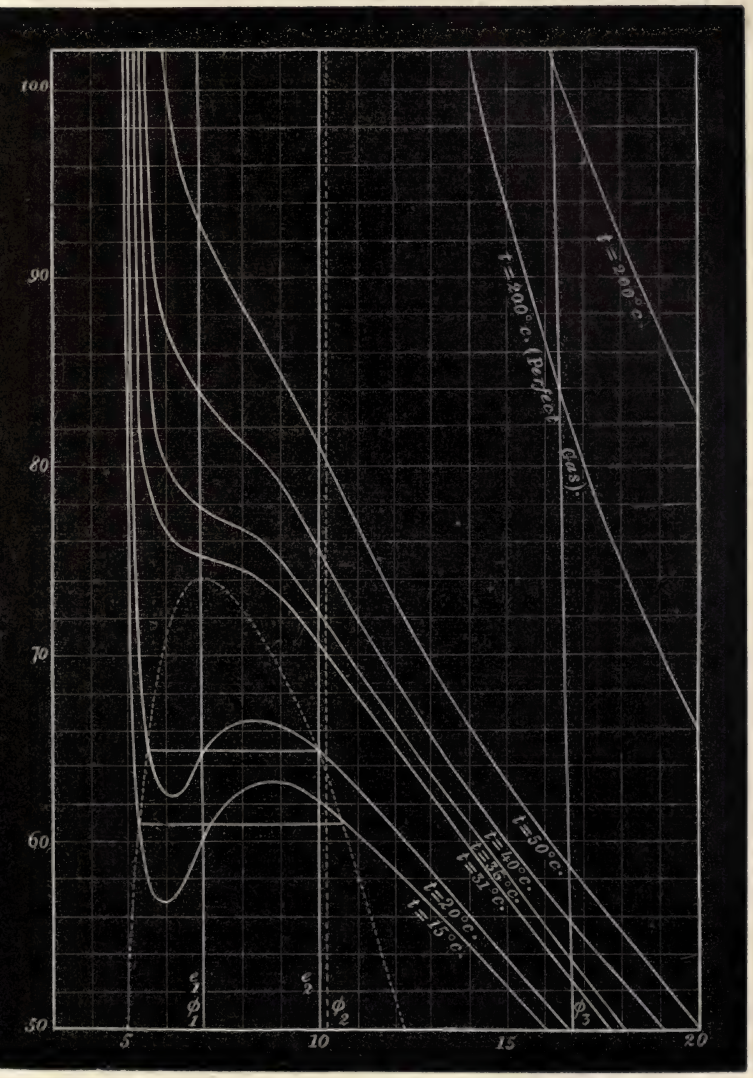

100
90
80
70
60
50
5
10
15
20
t = 200° c.
t = 200° c. (Perfect Gas).
t = 50° c.
t = 40° c.
t = 35° c.
t = 31° c.
t = 20° c.
t = 15° c.
ζ1
φ1
ζ2
φ2
φ3

FIG. 11.

part parallel to the volume-axis, but that it has a waved form, as shown in the diagram. Part of the waved portion represents an unstable state, since pressure and volume increase together. Thomson made this suggestion in order to avoid discontinuity in the curve.

The lines e_1, e_2 are lines of constant energy, and those marked ϕ_1, ϕ_2, ϕ_3 are lines ofconstant entropy.

25. It is only when the temperature considered happens to be one corresponding to a contour in the diagram that the relation of pressure and volume can be accurately found from the above diagram. This defect may be got rid of by the use of trilinear co-ordinates; and, in addition, the variation of a fourth quantity can be shown. In illustration of this we may take the case of a perfect gas, for which we have the equation

$$pv = Rt,$$

where p, v, and t represent respectively the pressure, volume, and

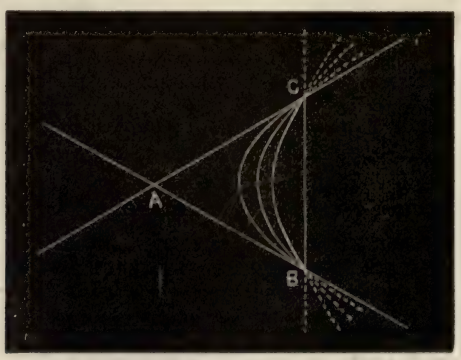


FIG. 12.

temperature, and R is a quantity which depends on the nature of the gas. The triangle of reference is made equilateral in Fig. 12. The ratios of the distances of a point from the vertical and the inclined sides of the triangle are the ratios of the temperature, pressure, and volume respectively. The contours for different values of R are shown in the figure, and the equation shows that they are hyperbolas, with vertical and horizontal axes. No part of the hyperbolas outside the triangle of reference has any physical meaning; for, in that case, pressure, volume, or temperature (or any two, or all, of them) would be negative, which is impossible in the case of a gas. Evidently pressure, volume, and temperature are continuously represented for any one gas.

Of course, when we wish to find the absolute values of the co-ordinate quantities, we must use the equation which connects them. This is not necessary when we use Cartesian co-ordinates.

The figure obviously represents the contours of a surface by planes parallel to the plane of the diagram, which may be looked upon as that corresponding to zero value of R. When R is zero, the hyperbola becomes a pair of straight lines which coincide with the sides AB, AC, of the triangle of reference. When R is infinite, the side BC is part of the corresponding hyperbola. All parts of lines through B and C perpendicular to the plane of the diagram lie upon the surface. These lines separate the parts of the surface which correspond to real physical states from those which do not. Outside the triangle the surface evidently overhangs the plane of the paper.

The value of R being given, let P (Fig. 13) be the point which gives

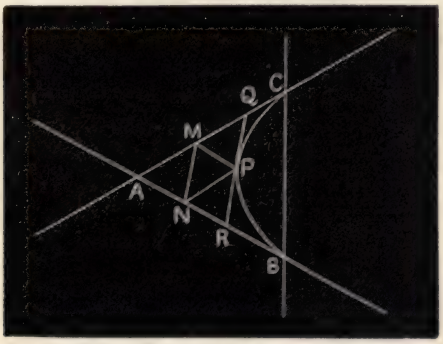


FIG. 13.

the proper ratios of $p, v,$ and t. Draw PM, PN, parallel to the sides of the triangle. Since the asymptotes of the hyperbola are parallel to these sides, it follows that that part of the tangent at P, which is intercepted by the sides of the triangle, is bisected at the point of contact. Therefore $AM = MQ$, and $AN = NR$. Now the compressibility k of a gas is given by the ratio dv/vdp, where dv is the small alteration of volume produced by the small change of pressure dp. But $dv/dp = MQ/MP = NP/MP = v/p$. Hence $k = 1/p$, that is, the compressibility of a perfect gas is the reciprocal of the pressure. Similarly it can be shown that the expansibility is proportional to the absolute temperature.

The work done during isothermal expansion can also be found from the diagram. The position of the point P gives the mutual

ratios of p, v, and t; but since t has a known constant value, the actual values of p and v are also known. Hence PN ($=p$ cosec BAC) is a known function of v. If P move to P', the area $PNN'P' = \int PNdv$ (§ 34) = cosec $BAC \int pdv$ is a known multiple of the work done.

26. The applicability of the method of contours to other physical problems is evident. Electric stream-lines and equipotential-lines may be regarded as contours of a surface, and the number of equipotential lines which cross unit length of a stream-line may be used to indicate the strength of the current. So also air-current lines and isobars, isothermals and flow-lines of heat, etc., are rectangular systems of contours.

CHAPTER IV.

VARYING QUANTITIES.

27. Though every quantity, whatever be its nature, has magnitude, no quantity can be said to be large or small *absolutely*. When we speak of the size of any body we mean its size relatively to the size of some other body with which we compare it. A yard is large if we compare it with an inch; it is small when compared with a mile. In the former case the number which represents it is more than 60,000 times larger than the number by which it is represented in the latter case. A mere number is therefore useless as regards the statement of magnitude, except when accompanied by a clear indication of what the thing measured is compared with. The quantity in terms of which the comparison is made is called the *unit*, and the number which tells how often this unit is contained in a given quantity is called the *numeric*.

All dynamical quantities may be made to depend upon *three* units only. These are the units of *mass* (quantity of matter), *length*, and *time*. Thus speed, being measured by the distance traversed in a certain time, depends upon the unit of length directly, and upon the unit of time inversely. Hence by doubling the unit of length we double the speed unit, and therefore halve the numeric of any given speed; whereas by doubling the unit of time we halve the speed unit, and therefore double the numeric of a given speed. Again, acceleration, being measured by the increase of speed in a certain time, depends upon the unit of speed directly and upon the unit of time inversely; that is, it depends directly upon the unit of length and inversely upon the square of the unit of time. The manner in which the fundamental units are involved in any quantity determines the *dimensions* of that quantity. If M L and T represent the units of mass, length, and time, the dimensions of speed and acceleration are indicated by the symbols $[LT^{-1}]$ and $[LT^{-2}]$ respectively, and the dimensions of energy (§ 7) by $[ML^2T^{-2}]$.

28. When two quantities are so related that any change in the numerical value of one of them is accompanied by a change in the numerical value of the other, each quantity is called a *function* of the other. If to one value of the first there corresponds one, and only one, value of the second, the second is called a *single-valued* function of the first; but if to a given value of the one there correspond more than one value of the other, the latter is said to be a *multiple-valued* function of the former. For example, x^2+2x-8 is a single-valued function of x; since, if we give x any value, there will be *one* corresponding value of x^2+2x-8. On the other hand, x is a double-valued function of x^2+2x-8; since, if we give any value to x^2+2x-8, we find that x has in general two distinct values. Again $\sin x$ is a single-valued function of x, while $\sin^{-1}x$ (the angle whose sine is x) is a multiple-valued function of x.

The relation between the two quantities can be expressed by means of an analytical equation or by means of a curve, as is indicated in § 17. The general expression of the relation may be given in the form

$$y = f(x),$$

where the quantity y is regarded as being dependent upon x, and the equation simply reads 'y is some function of x.' The function denoted by f may be of various kinds—it may be *algebraical* (*e.g.*, $y = ax+bx^2$), *trigonometrical* (*e.g.*, $y = \sin x+\cos x$), *exponential* (*e.g.*, $y=a^x$), etc.; and under each kind there may be a number of different *forms*, *e.g.*, a 'series of powers,' a 'product of tangents,' etc.

In the above formula x is supposed to be that quantity the value of which is arbitrarily varied. It is therefore called the *independent variable*, while y is termed the *dependent variable*.

29. If y varies in value uniformly when x varies uniformly, the quotient of any increment of y by the simultaneous increment of x is constant, however large or small either of the increments may be. For the condition means that y is proportional to x, or to x + a constant, say $y=kx+c$ where k and c are constants. Hence if y changes from y_1 to y_2 when x changes in value from x_1 to x_2 we have $y_1=kx_1+c$, $y_2=kx_2+c$, and therefore $y_2-y_1=k\ (x_2-x_1)$, which proves the above statement, since x_2-x_1 may have any value we please. If the values of y and x be represented as the ordinates and abscissæ respectively of a curve, we see that the constant c is represented by the length OA (Fig. 14) (since it is the value of y when $x=0$), that simultaneous values of x and y are represented by points on the line AB, and that k is represented by the tangent of the inclination of that line to the x-axis; hence, by this graphical process

also, the truth of the statement is evident. When they are thus related, y and x are said to be *linear* functions of each other.

But when y is not a linear function of x, it is evident that the ratio of their simultaneous increments depends upon the absolute values of these increments. In the former case that ratio denoted the *rate* at which y varied when x varied; but, in the case now supposed, it does not in general give the true value of that rate. It is of extreme importance that we should have a means of finding the true rate of variation whatever be the nature of the relation connecting the two quantities. Let $A'B'$ (Fig. 14) be the curve which represents y as a function of x, and let P be the point at which we wish to find the rate of variation of y when x alters. Take another point P' on the curve, and let x' and y' be the values of its co-ordinates. It is obvious that in general the ratio of $y'-y$ to $x'-x$ does not give the true rate. It gives instead the rate of

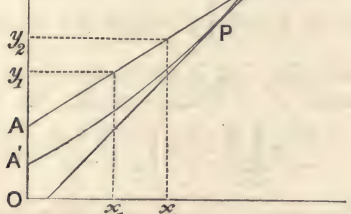


FIG. 14.

variation corresponding to the line joining P and P'. But the line PP' coincides more and more nearly with the curve as P' approaches to P; and we can take P' closer to P than any assigned finite quantity however small, so that the difference in direction between the line and the curve can be made smaller than any assignable angle. Ultimately—when P' is infinitely near to P—the rates of variation corresponding to the curve and the line are identical. Hence, geometrically, the true rate at any point is given by the tangent of the angle which the line touching the curve at that point makes with the x-axis; or, analytically, it is given by the ratio of $y'-y$ to $x'-x$ when $x'-x$ is made indefinitely small.

This method is only applicable strictly to the case of quantities which have no *sudden* change in their rate of variation. But, in

any such case, we can apply the method up to the point at which the sudden change occurs, and also, separately, beyond this point: and this is all that is required.

It is usual to write dy and dx instead of $y'-y$ and $x'-x$ when these quantities are infinitely small, so that the symbol d means the infinitely small increment (or 'differential,' as it is usually termed) of the quantity to which it is prefixed. In general, dy/dx is a function of x; so that, if the original relation is $y=f(x)$, it is usual to represent the quantity dy/dx by the symbol $f'(x)$, and to call it the first *derived function* of y with respect to x. We may deal similarly with the relation $y=f'(x)$, and obtain the second derived function which is indicated by $f''(x)$, and so on. It must be carefully observed that dx and dy, being quantities of the same kind as x and y, are subject to identically the same laws.

30. We shall now find the rates of variation of certain functions which will be of use subsequently; first of all, of *rational algebraical functions*.

$$(1.)\quad y = ax + b,$$

a and b being constants. We shall use the ordinary sign for a limit, according to which $\mathop{\mathrm{L}}_{x'=x}$ denotes the limiting value of the quantity to which the sign is prefixed when in that quantity we put $x'=x$, *i.e.*, make $x'-x$ infinitely small. We have

$$\frac{dy}{dx} = \mathop{\mathrm{L}}_{x'=x} \frac{y'-y}{x'-x} = \mathop{\mathrm{L}}_{x'=x} \frac{a(x'-x)}{(x'-x)} = a,$$

a result which has already been given in the preceding section.

This example shows that a constant term in a function does not appear in its derived function.

$$(2.)\quad y = ax^2$$

This gives

$$\frac{dy}{dx} = \mathop{\mathrm{L}}_{x'=x} \frac{a(x'^2-x^2)}{x'-x} = \mathop{\mathrm{L}}_{x'=x} \frac{a(x'+x)(x'-x)}{x'-x} = 2ax.$$

$$(3)\quad y = ax^n,$$

gives
$$\frac{dy}{dx} = \mathop{\mathrm{L}}_{x'=x} \frac{a(x'^n-x^n)}{x'-x}$$
$$= a \mathop{\mathrm{L}}_{x'=x} \frac{(x'-x)\,(x'^{n-1}+x'^{n-2}x+\ldots x'x^{n-2}+x^{n-1})}{x'-x}$$
$$= anx^{n-1}.$$

$$(4.)\quad y = ax^{\frac{m}{n}}.$$

This gives

$$y^n = a^n\ x^m.$$

$$\frac{d(y^n)}{dx} = \frac{d(a^n x^m)}{dx}$$

$$\frac{d(y^n)}{dy} \cdot \frac{dy}{dx} = a^n \frac{d(x^m)}{dx}$$

$$ny^{n-1}\frac{dy}{dx} = a^n m x^{m-1} \quad \ldots\ldots\ldots\ldots \text{ by (3)}$$

$$\frac{dy}{dx} = a^n \frac{m}{n} x^{m-1} \frac{1}{y^{n-1}}$$

$$= a\frac{m}{n} \frac{x^{m-1}}{x^{m\frac{n-1}{n}}} = a\frac{m}{n} x^{m-1-m+\frac{m}{n}} = a\frac{m}{n} x^{\frac{m}{n}-1},$$

We see from (3) and (4) that when y is given as a positive power of x (*whole or fractional*), its x-rate of variation is got by the following rule: *Multiply by the index of* x *and lower the index by unity.* Example (7) will show that the rule is not restricted to positive powers of x, but holds also for negative powers.

$$(5.)\quad y = ax^n + bx^m$$

$$\frac{dy}{dx} = \mathop{\mathrm{L}}_{x'=x}\left[a\frac{(x'^n - x^n)}{x' - x} + b\frac{(x'^m - x^m)}{x' - x}\right] = \mathop{\mathrm{L}}_{x'=x}\left[a(x^{m-1} + x'^{n-2}x + \ldots) + b(x'^{m-1} + x'^{m-2}x + \ldots\ldots)\right]$$

$$= anx^{n-1} + bmx^{m-1}.$$

From this example we see that *the rate of variation of a sum of such terms is the sum of the rates of variation of each.*

$$(6.)\quad y = uv,$$

where u and v are functions of x. As before

$$\frac{dy}{dx} = \mathop{\mathrm{L}}_{x'=x}\frac{y' - y}{x' - x},$$

and, just as y increases by the quantity dy when x increases by dx, u and v will increase by du and dv. Therefore

$$\frac{dy}{dx} = \frac{(u+du)\ (v+dv) - uv}{dx} = \frac{udv + vdu + dudv}{dx}.$$

But the third term in the numerator of this fraction, being the *product* of infinitesimally small quantities, is infinitely smaller than the others, and so vanishes in comparison with them. Hence

$$\frac{dy}{dx}=\frac{d(uv)}{dx}=u\frac{dv}{dx}+v\frac{du}{dx}.$$

That is to say, *the rate of variation of a product of two quantities is obtained by taking the sum of the products of each quantity into the rate of variation of the other.* As an illustration we may find the x-rate of variation of $x^m \cdot x^n$. Here the result is

$$x^m \cdot nx^{n-1}+mx^{m-1} \cdot x^n=(m+n)x^{m+n-1},$$

which agrees with the result of (3).

If, in the above example, y is a constant, *i.e.*, if the product of u and v is constant, we get $udv=-vdu$.

$$(7.)\quad y=\frac{u}{v}.$$

Let $v=\frac{1}{w}$ and therefore $y=uw$, and we get

$$\frac{dy}{dx}=u\frac{dw}{dx}+w\frac{du}{dx}$$

But, as has just been remarked, we have $vdw=-wdv$, since $wv=1$, and so

$$\frac{dy}{dx}=w\frac{du}{dx}-\frac{uw}{v}\frac{dv}{dx}=\frac{1}{v}\frac{du}{dx}-\frac{u}{v^2}\cdot\frac{dv}{dx}$$

$$=\frac{v\frac{du}{dx}-u\frac{dv}{dx}}{v^2},$$

which gives the rule for the rate of variation of a quotient of two functions.

If $u=x^n$, $v=x^m$, we get

$$\frac{dy}{dx}=\frac{x^m \cdot nx^{n-1}-x^n \cdot mx^{m-1}}{x^{2m}}=(n-m)x^{n-m-1}.$$

Again, if u is constant, we have

$$\frac{dy}{dx}=-\frac{u}{v^2}\cdot\frac{dv}{dx}.$$

Thus, if $y=ax^{-n}$, a takes the place of u, and x^n replaces v. Therefore

$$\frac{dy}{dx}=-\frac{a}{x^{2n}}\cdot\frac{dx^n}{dx}=-\frac{na}{x^{2n}}x^{n-1}=-nax^{-n-1},$$

which proves the applicability, to the case of negative indices, of the rule given under example (4).

In many cases y will be given, not in terms of x as hitherto assumed, but in terms of some function of x, say u. The ordinary principles of algebra then show (see the remark at the end of last section) that

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

For example, $y = (a^2+x^2)^{\frac{1}{2}} = u^{\frac{1}{2}}$

$$\frac{dy}{dx} = \frac{d(u^{\frac{1}{2}})}{dx} = \frac{du^{\frac{1}{2}}}{du}\frac{du}{dx} =$$

$$\tfrac{1}{2}u^{-\frac{1}{2}}\frac{du}{dx} = \tfrac{1}{2}u^{-\frac{1}{2}}\frac{d(a^2+x^2)}{dx} = \tfrac{1}{2}u^{-\frac{1}{2}}\frac{dx^2}{dx}$$

$$= u^{-\frac{1}{2}}x = (a^2+x^2)^{-\frac{1}{2}}x.$$

31. We shall now consider such rates of variation of *trigonometrical* functions as may be of use later.

(1) $y = \sin x.$

$$\frac{dy}{dx} = \mathop{\mathrm{L}}_{x'=x} \frac{\sin x' - \sin x}{x'-x} = 2\mathop{\mathrm{L}}_{x'=x} \frac{\sin\left(\frac{x'}{2} - \frac{x}{2}\right)\cos\left(\frac{x'}{2} + \frac{x}{2}\right)}{x'-x}$$

But, when $x' = x$, $\sin(x' - x) = x' - x$, and therefore the rate of variation is $\cos x$.

It follows at once from this result, by writing $x + \frac{\pi}{2}$ instead of x, that if

(2) $y = \cos x$

$$\frac{dy}{dx} = -\sin x.$$

Again, suppose

(3) $y = \sec x.$

Let $\cos x = u$, and we get

$$\frac{dy}{dx} = \frac{du^{-1}}{du}\frac{du}{dx} = -u^{-2}\frac{du}{dx} = -\sec^2 x \cdot \frac{du}{dx}$$

$$= \sec^2 x \cdot \sin x = \sec x \tan x.$$

From this result we readily obtain, when

(4) $y = \operatorname{cosec} x,$

$$\frac{dy}{dx} = -\operatorname{cosec} x \cot x.$$

By means of the result of example (7) in § 30 we can find the derived function of

$$(5)\quad y = \tan x = \frac{\sin x}{\cos x}.$$

That result gives

$$\frac{dy}{dx} = \frac{\cos x \frac{d \sin x}{dx} - \sin x \frac{d \cos x}{dx}}{\cos^2 x}$$

$$= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x\,;$$

and, as formerly, by the substitution of $x + \frac{\pi}{2}$ for x, we deduce, from

$$(6)\quad y = \cot x,$$

$$\frac{dy}{dx} = -\operatorname{cosec}^2 x.$$

32. It will be instructive to deduce the results, which we have just obtained, by a geometrical process. Take two near points, P and P′, on the circumference of the circle APB. Denote POA by θ and POP′ by $d\theta$, and suppose that the radius of the circle is

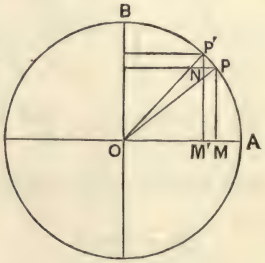


FIG. 15.

unity. Let $OM = x$ and $MP = y$; then $NP' = dy$ and $-NP = dx$. The angle $d\theta$ is assumed to be infinitesimally small (so that the arc PP′ is practically a straight line), and it is measured by the ratio of PP′ to OP; that is, by PP′, since OP is unity. Also $NP'P = \theta$. Hence $NP'/PP' = \cos\theta = dy/d\theta$, and $-NP/PP' = -\sin\theta = dx/d\theta$. But $x = \cos\theta$, $y = \sin\theta$; and therefore

$$\frac{d \sin\theta}{d\theta} = \cos\theta, \quad \frac{d \cos\theta}{d\theta} = -\sin\theta.$$

Also $$\sec\theta = 1/x,$$

and therefore

$$d \sec\theta = \frac{1}{x + dx} - \frac{1}{x} = \frac{-dx}{x(x + dx)} = -\frac{dx}{x^2}$$

ultimately. But

$$-\frac{dx}{x^2}=\frac{1}{x^2}\left(-\frac{dx}{d\theta}\right)d\theta=\frac{1}{x^2}\sin\theta\,d\theta=\sec^2\theta\sin\theta\,d\theta=\sec\theta\tan\theta\,d\theta.$$

Similarly,

$$d\tan\theta=\frac{y+dy}{x+dx}-\frac{y}{x}=\frac{xy+xdy-xy-ydx}{x(x+dx)}=\frac{xdy-ydx}{x^2}$$

$$=\frac{\cos\theta\,.\,\cos\theta\,d\theta-\sin\theta\,(-\sin\theta)\,d\theta}{\cos^2\theta}=(1+\tan^2\theta)\,d\theta=\sec^2\theta\,d\theta.$$

33. Lastly we may obtain the rate of variation of the exponential and logarithm of a varying quantity.

(1) $y=e^x$

$$\frac{dy}{dx}=\mathop{\mathrm{L}}_{x'=x}\frac{e^{x'}-e^x}{x'-x}=\mathop{\mathrm{L}}_{x'=x}\frac{e^x(e^{x'-x}-1)}{x'-x}.$$

But e, which is the base of the Napierian Logarithms, is by definition the limiting value of $(1+u)^{\frac{1}{u}}$ when u is a vanishingly small quantity; so that we may write

$$e=\mathop{\mathrm{L}}_{x'=x}(1+x'-x)^{\frac{1}{x'-x}}$$

Therefore the limiting value of $e^{x'-x}$ is

$$\mathop{\mathrm{L}}_{x'=x}\left[\mathop{\mathrm{L}}_{x'=x}(1+x'-x)^{\frac{1}{x'-x}}\right]^{x'-x}=\mathop{\mathrm{L}}_{x'=x}(1+x'-x);$$

and therefore

$$\frac{dy}{dx}=\mathop{\mathrm{L}}_{x'=x}e^x\,\frac{(1+x'-x-1)}{x'-x}=e^x.$$

By means of this result we can deduce the derived function of

(2) $y=\log x.$

For this gives $x=e^y$, and therefore $dx/dy=e^y$, so that

$$\frac{dy}{dx}=e^{-y}=e^{-\log x}=e^{\log x^{-1}}=\frac{1}{x}.$$

Inverse Problem.

34. In the immediately preceding sections we have dealt with the problem—to find the rate of variation of a quantity which is a given function of some independent variable. The inverse problem—to find the value of a quantity, the rate of variation of which is given—is of quite as great importance in physical inquiries; and, in order to solve it, we have merely to reverse the former process.

In illustration we may consider the simplest possible case in which we have $y=x$. Here dy/dx is constant and numerically equal to unity (§ 30, example (1)). The relation $y=x$ is indicated in Fig. 16 by the line bisecting the angle xoy, and the values of dy/dx are

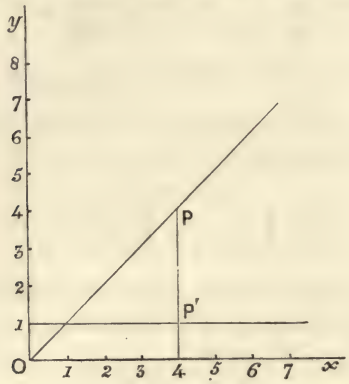


FIG. 16.

given by points on the horizontal straight line passing through the point $y=1$. If we choose any point P on the line $y=x$, and draw through it a line parallel to the y-axis cutting the line indicating the value of dy/dx in P', it is evident from the figure that the area of the

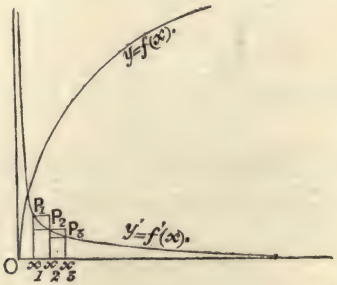


FIG. 17.

rectangle OP' is numerically equal to the value of y at P. For the ordinate of P is four units in length, and OP' contains four square units.

Now take any curve (the particular curve shown in Fig. 17 is the

quadrant of a circle) representing graphically the relation $y=f(x)$, and draw another curve representing the relation $y'=f'(x)$. This may be called the 'derived' curve of the former, since its ordinates give the values of dy/dx at the corresponding points, *i.e.*, they give the values of the derived function. Take three near points, $P_1P_2P_3$, on the derived curve. The area included between the curve and the axes is greater than the sum of such areas as P_2x_1, but is less than the sum of such areas as P_1x_2. These sums differ by the sum of the rectangles P_1P_2, etc. But, as x_2-x_1, etc., are made less and less and ultimately vanish, these rectangles become smaller and smaller and finally vanish. Of course any area P_1x_2 becomes infinitesimally small as x_2 approximates to coincidence with x_1; but P_1P_2 becomes infinitely small in comparison with P_1x_2—infinitely small though it be. The magnitude of each little area similar to P_1x_2 is $y'dx$, where dx represents as before an infinitesimally small increment of x. But $y'=dy/dx$, and therefore $y'dx=dy$. So that, finally, the limiting value of the sum of the rectangles (the number of which increases indefinitely as their magnitude diminishes without limit) up to a given value of x, is numerically equal to the value of y, at that given value of x, in the original curve. This limit to which the sum approaches is indicated by the symbol

$$\int y'dx,$$

and the quantity itself is called the *integral* of y' with respect to x. That is to say, while y' is called the 'derived function' of y with respect to x, y is said to be the 'integral' of y' with respect to the same variable quantity. If it were necessary to preserve similarity of nomenclature, we might term y the 'primitive function' of y'. Analytically, the operation indicated by $\int$ is an operation which undoes the effect of the operation indicated by the symbol d. For

$$y=\int y'dx=\int\frac{dy}{dx}dx=\int dy.$$

The symbol d signifies an infinitesimal difference; the symbol $\int$ signifies the sum of an infinite number of infinitesimally small differences. Indeed the symbol of integration is merely an exaggerated form of the letter S, denoting a sum.

In order to find the integral of y' we have to answer the question, What function has y' as its derived function? A considerable knowledge of derived functions is therefore essential.

35. We shall now consider some useful examples, and shall take them in an order similar to that which was observed when finding the derived functions.

(1) Evaluate $y = \int a dx$.

We have merely to write down that function the derived function of which with respect to x is a. In the direct process we had to lower the power of x by unity and multiply by the undiminished power as a factor. Therefore, in the inverse process, we must *raise* the power by unity and *divide* by the *increased* power. But, in the above example, we may assume x^0 as a factor since its value is unity; so that ax^0 is the quantity of which we have to obtain the integral. In accordance with the rule just given the result is ax. But we must remember that in the direct process all constant terms disappear (§ 30, example (1)); so, to the result as above obtained, we must always add on a constant. Thus:

$$y = \int a dx = ax + b.$$

The constant b is quite arbitrary unless some condition is laid down which determines it. Thus the condition might be that when x is unity, y is equal to $a+3$, in which case we see that the value of b is 3.

(2) Evaluate $y = \int 2ax dx$.

In accordance with the rule which requires us to increase the power of x by unity and divide by the power so increased and add a constant, we get

$$y = 2a\tfrac{1}{2}x^2 + b = ax^2 + b.$$

Similarly from

$$(3)\quad y' = anx^{n-1}$$

we deduce

$$y = ax^n + b.$$

Or, changing n into $n+1$, from

$$y' = ax^n$$

we obtain

$$y = a\frac{1}{n+1}x^{n+1} + b,$$

which is true whether n be positive or negative, whole or fractional.

The proof, given in § 30, that the rate of variation of a sum of powers of the independently varying quantity is the sum of the rates of variation of each term, gives at once, by inversion, the rule that *the integral of a sum of powers is the sum of the integrals of each term.* From

$$(4)\quad y' = ax^n + bx^m + \ldots$$

we can at once write down

$$y = a\frac{x^{n+1}}{n+1} + b\frac{x^{m+1}}{m+1} + \ldots.$$

In example (6), § 30, it has been shown that, when $y = uv$, the

increment of y consequent upon alteration of the quantity of which u and v are functions is

$$dy = udv + vdu.$$

Hence, from

$$(5)\quad y' = uv' + vu',$$

we obtain

$$y = uv + a \text{ constant}.$$

Also from example (7) of the same paragraph we see that

$$(6)\quad y' = \frac{vu' - uv'}{v^2}$$

gives

$$y = \frac{u}{v} + a \text{ constant}.$$

36. In section 34 we found how we might represent graphically the value of y for different values of x by means of an area included between the x-axis, the curve representing y' as a function of x, and two ordinates of that curve. By means of this method we can obtain an independent proof of the result of example (3) above. Let the curve in the diagram represent the relation $y' = x^n$. The value of y corresponding to x is represented by the area OxP; and

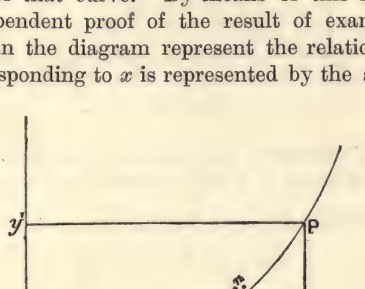


FIG. 18.

$OxP = OxPy' - OPy'$. Now, in precisely the same way that OxP represents $\int y'dx$, the area OPy' represents the quantity $\int xdy'$, as is indicated by the horizontal rectangle. But $\int xdy' = \int xd\, x^n = \int x \,.\, \frac{dx^n}{dx} dx = \int x \,.\, nx^{n-1} \,.\, dx = n\int x^n dx$. Also $OxP = \int y'dx = \int x^n dx$, and $OxPy' = xy' = x^{n+1}$. Therefore $x^{n+1} = \int x^n dx + n\int x^n dx = (n+1)\int x^n dx = (n+1)\int y'dx = (n+1)y$; that is

$$y = \frac{x^{n+1}}{n+1}.$$

In Fig. 19 the rectangular area OP represents the product of u and v. When u and v increase or decrease simultaneously, the increase of uv is evidently $udv + vdu$ (the little rectangle at P being neglected). And also $uv = \int vdu + \int udv$, which makes the result of example (5), § 35, almost self-evident. If u decreases as v increases (as in passing from P'_1 to P'_2) the area uv increases by the quantity udv, but *decreases* by the amount vdu. But, in this case, du is itself negative; so that the result is still $d(uv) = udv + vdu$.

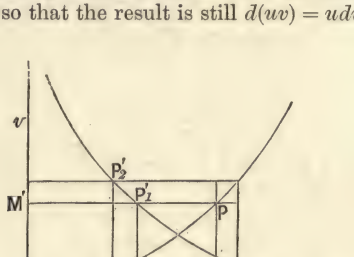


FIG. 19.

If the curve P represents the value of v in terms of u, we may take its reciprocal curve P' which represents w in terms of u, where $w = \frac{1}{v}$. The diagram gives (as above)

$$uw = \int udw + \int wdu.$$

But when any quantity v increases by the amount dv, its reciprocal decreases by $\frac{1}{v} - \frac{1}{v+dv} = \frac{dv}{v(v+dv)} = \frac{dv}{v^2}$ ultimately; that is, the increase is $-\frac{dv}{v^2}$.

Therefore

$$\frac{u}{v} = \int \frac{du}{v} - \int \frac{udv}{v^2} = \int \frac{vdu - udv}{v^2},$$

which is the result of example (6) preceding.

37. From the results of examples (1) and (2) of § 31, we can at once write down

$$\int \sin x dx = -\cos x \text{ and } \int \cos x dx = \sin x,$$

because the derived functions of $\cos x$ and $\sin x$ are respectively $-\sin x$ and $\cos x$.

Since we know that $d \tan \theta / d\theta = \sec^2 \theta$, we can put

$$\tan \theta = \int \sec^2 \theta \, d\theta;$$

and similarly

$$\int \sec \theta \tan \theta \, d\theta = \int \sec^2 \theta \sin \theta \, d\theta = \sec \theta$$

$$\int \operatorname{cosec} \theta \cot \theta \, d\theta = \int \operatorname{cosec}^2 \theta \cos \theta \, d\theta = -\operatorname{cosec} \theta,$$

and

$$\int \operatorname{cosec}^2 \theta \, d\theta = -\cot \theta.$$

38. Lastly, since (§ 33) $de^x/dx = e^x$ and $d \log x/dx = \frac{1}{x}$, we get

$$\int e^x dx = e^x, \text{ and } \int \frac{dx}{x} = \log x.$$

CHAPTER V.

MOTION.

39. *Position.* — The *position* of a point in space is completely determined when *three* independent conditions are given, each of which it satisfies. And its position can only be given *relatively* to that of another point, for we do not know any point of which we can assert that it is absolutely fixed. We may say that one point is so much to the north or south of another, so far to the east or west of it, and so much higher or lower; or we may say that it is so far distant from the other, that the line joining the two is inclined at a certain angle to the vertical, and that the vertical plane through the two has a given inclination to the vertical north-and-south plane. The given quantities which determine the position are called the *co-ordinates* of the point. The first case furnishes an example of the ordinary *Cartesian* system of rectangular co-ordinates, the second illustrates the system known as *polar* co-ordinates. In the Cartesian system

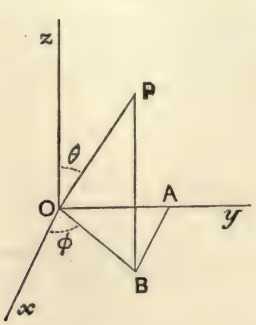


FIG. 20.

the co-ordinates are usually denoted by the letters x, y, z; in the polar system, the length is generally denoted by r, and the angles

by θ and ϕ. Thus, in Fig. 20, the relative position of P to O may be given by $x=\mathrm{AB}$, $y=\mathrm{OA}$, $z=\mathrm{BP}$; or by $r=OP$, $POz=$, $xOB=\phi$.

40. *Displacement.*—When two points occupy different positions, we speak of the *displacement* of the one from the other; and it is by means of the displacement that we determine their relative position. Two ideas are essentially involved in the term—the idea of *magnitude* and the idea of *direction*. If we know only that one point is distant three feet from another, we cannot tell what position it occupies on the spherical surface all points of which satisfy this condition. Other two conditions are required to fix the direction also, that is, to determine the displacement.

Addition of two displacements is effected when we find the single displacement, which produces the same result as the two do when

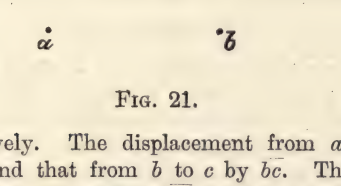


FIG. 21.

applied consecutively. The displacement from a to b may be denoted by $\overline{ab}$, and that from b to c by $\overline{bc}$. Then $\overline{ab}+\overline{bc}=\overline{ac}$. But the displacement denoted by $\overline{bc}$ might have been performed first. The effect would be to transfer a to b', which is a point such that $\overline{ab'}$ is equal and parallel to $\overline{bc}$. This follows since a displacement does not involve the idea of position, but only the ideas of magnitude and direction; in fact, $\overline{ab'}$ is the *same* displacement as $\overline{bc}$. And, similarly, $\overline{b'c}$ is the same as $\overline{ab}$; so that $\overline{ab'}+\overline{b'c}=\overline{bc}+\overline{ab}$ $=\overline{ac}=\overline{ab}+\overline{bc}$. Also, since a displacement is reversed when its direction is reversed, we have $\overline{ab}=-\overline{ba}$; and the ordinary laws of algebra apply to addition and subtraction of displacements. The *lines* ab, bc, etc., may be used to represent the *displacements* $\overline{ab}$, $\overline{bc}$, etc.; for a line involves necessarily, and only, magnitude and direction.

A line, given in magnitude and direction, is called a *vector;* and any quantity which, like a displacement, requires for its complete representation a directed line is called a *vector quantity*. In the course of this chapter, and of Chap. VI., we shall get various examples of such quantities.

A quantity which is independent of direction — which merely possesses magnitude—is called a *scalar* quantity, that is, it is completely determined by measurement on a suitable scale. When treating such quantities algebraically it is usual to denote scalars by ordinary letters, as a, b, x, y, etc., and vectors by Greek letters as α, β, γ, etc.

Any displacement in space can be fully represented in terms of three distinct unit vectors and three independent scalars. Thus, in the figure of last section, α may represent a unit length drawn in the direction AB, and the line AB may contain x units of length, so that the vector $\overline{AB}$ is $x\alpha$. Similarly, if β is a unit vector in the direction Oy, the vector $\overline{OA}$ may be denoted by $y\beta$; and the vector $\overline{BP}$ may be represented as $z\gamma$. Hence the vector $\overline{OP}$ is $x\alpha+y\beta+z\gamma$. But this quantity denotes merely the position of P *relatively* to that of O: and, consequently, if any other point P' is situated relatively to another point O' in the same way in which P is situated with respect to O, the vector $x\alpha+y\beta+z\gamma$ represents the displacement $\overline{O'P'}$.

If all the quantities x, y, z, are variable arbitrarily, the vector

$$\rho = x\alpha + y\beta + z\gamma$$

is the vector of *any* point in space. If x and y alone are variable arbitrarily, the point lies upon a plane parallel to the directions indicated by α, β. If one of the quantities alone varies, say x, the point lies on a line whose direction is indicated by α. If all three are fixed, the point is fully determined. More generally, if x, y, and z are connected by one, two, or three relations respectively, the equation indicates respectively a surface, a curve, or a definite point.

We have spoken of the displacement of one point relatively to another as that which determines the relative position in space of the first with respect to the second. Sometimes the two points may lie on a given surface or a given curve, and it is then frequently convenient to speak of the displacement *on the surface* or *along the curve*. This means that the magnitude of the displacement is to be measured *along* the surface or curve.

41. *Speed and Velocity.*—The displacement of a point may be constant, or it may vary. If it varies, we say that the point is in motion: so that by *motion* we mean change of relative position. [The science which treats of motion, and which is generally called *Kinematics*, therefore deals with the ideas both of space and of time.] The motion of a point is essentially a *translation*, for it has no separate parts which can rotate relatively to each other. Its position, we have seen, is not fixed unless three independent conditions

are given. The removal of one restraining condition leaves the point more free to move than before; and a point, the motion of which is unrestrained, is said to have three degrees of translational freedom.

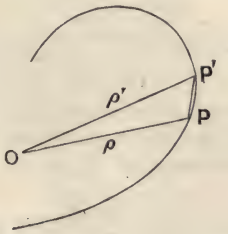


FIG. 22.

If P moves to P' (Fig. 22), the change of displacement is represented by the line PP', or (§ 40), by $\rho' - \rho$, where ρ' and ρ represent the vectors $\overline{OP'}$ and $\overline{OP}$. If this change occurs in time $t' - t$, the time-rate of change is $\lim_{t'=t} \frac{\rho' - \rho}{t' - t}$ or $\frac{d\rho}{dt}$. It is convenient to represent this quantity by $\dot{\rho}$; so that $\dot{\rho}$ is the time-rate of variation of ρ, or its first derived function with respect to the time as the independent variable (§ 28).

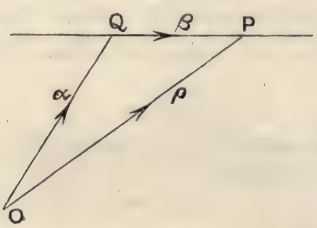


FIG. 23.

The mere *magnitude* of this time-rate is called the *Speed* of the moving point; but, when the *direction* is considered also, the term *Velocity* is used.

In illustration of this we shall consider the case of uniform motion in a straight line. Let QP (Fig. 23) be the line. Let Q be a fixed point on it, and let P be the uniformly moving point. Since P moves uniformly, the length of QP is proportional to the

time t reckoned from the instant at which P occupied the position Q; say QP $= at$, where a is a constant. Let β be a unit vector in the direction QP, and let α, ρ be the vectors from O to Q and P respectively. We have then

$$\rho = \alpha + at\beta.$$

Hence, by the principles of Chapter IV.,

$$\dot{\rho} = a\beta.$$

Here a is the *speed* of motion, and $\dot{\rho}$ is the *velocity*. The magnitude a is constant, and the direction β is also fixed. Also $\rho - \alpha = at\beta$. Hence the distance traversed is

$$s = at.$$

[This necessitates our defining unit speed as that speed with which unit distance is described in unit time.]

42. *Acceleration of Speed.* When the velocity of a moving point varies it is said to be *accelerated*, and the time-rate of its change is called the *acceleration*. Meantime we shall suppose that the change affects the magnitude only and not the direction: that is, we shall investigate non-uniform motion in a straight line. We may write (using the diagram of last section),

$$\rho = \alpha + x\beta,$$

where x is some function of t, and we get

$$\dot{\rho} = \dot{x}\beta.$$

Here $\dot{x}$ is the variable speed of motion. Forming the second derived function of ρ with respect to the time and denoting it by $\ddot{\rho}$, we obtain

$$\ddot{\rho} = \ddot{x}\beta.$$

If we assume the acceleration to be constant, that is, $\ddot{x} =$ a constant $= b$ (say), this becomes

$$\ddot{\rho} = b\beta.$$

But, § 35, this equation gives

$$\dot{\rho} = (a + bt)\,\beta + \gamma = \dot{x}\beta + \gamma$$

where a is the value of the speed when $t = 0$ (usually called the initial speed), and γ is a constant vector which vanishes if we suppose the velocity to have the direction β. And from this we deduce further

$$\rho = \beta\int \dot{x}dt = \beta\int (a+bt)\,dt = (c + at + \tfrac{1}{2}bt^2)\,\beta + \alpha = x\beta + \alpha.$$

If we suppose the point O to be in the line of motion, a will vanish, x will be the distance traversed from Q, and c will be the distance

from O to Q. In other words, c is the numerical value of ρ (called the *tensor* of ρ, and denoted by the symbol Tρ) when $t=o$. Thus under uniform acceleration in the direction of motion we get

$$\ddot{x}=b$$
$$\dot{x}=a+bt$$
$$x=c+at+\tfrac{1}{2}bt^2.$$

If the acceleration is negative, *i.e.*, opposite to the direction of motion, we must prefix the minus sign to the quantity b. If the motion begins from rest, the quantities a and c vanish.

As a special case, when a body falls from rest under the action of gravity, in which case the acceleration is denoted by the letter g, we get

$$\ddot{x}=g,\ \dot{x}=gt,\ x=\tfrac{1}{2}gt^2.$$

Again, if the body is thrown upwards with speed V and we consider the upward direction to be positive, the equations become

$$\ddot{x}=-g,\ \dot{x}=\mathrm{V}-gt,\ x=\mathrm{V}t-\tfrac{1}{2}gt^2.$$

The second of these equations enables us to tell at once how long the body will take to rise to its greatest height above the ground. For, when the body is at its greatest height, the speed vanishes, that is, $\dot{x}=o$, and therefore

$$t=\frac{\mathrm{V}}{g}.$$

From the third equation we can tell after what time it will reach the ground. Since we are supposing x to be measured from the surface of the earth, the condition is $x=o$. This condition gives either $t=o$, or

$$t=\frac{2\mathrm{V}}{g}.$$

Hence the body takes as long a time to fall from its greatest height as it took to rise to it. Again, we have $\dot{x}^2=\mathrm{V}^2-2\mathrm{V}gt+g^2t^2=\mathrm{V}^2-2g(\mathrm{V}t-\tfrac{1}{2}gt^2)=\mathrm{V}^2-2gx$. When the point is at its greatest height, $\dot{x}=o$, and

$$x=\frac{\mathrm{V}^2}{2g}.$$

Also, when $x=o$, we get $\dot{x}=\pm\mathrm{V}$; that is, the point reaches the ground with speed $-\mathrm{V}$ equal and opposite to the speed of projection.

43. *Curvature. Acceleration of Velocity.* — In the preceding section we have supposed the direction of motion to be unaltered. When the direction changes the path of the moving point is said to

be 'curved.' The tangent to the curve gives the direction of motion at any instant, and the limiting value of the ratio of the angle between two tangents at near points to the length of path between these points as they are taken closer and closer together and finally coincide is called the *Curvature* of the path at that place. Thus the curvature is the rate of change of the direction of motion per unit length of the curve. If the tangent to the curve at any point makes an angle θ with any fixed line in its plane, while s is the length of the curve, this definition gives as the measure of the curvature the quantity $d\theta/ds$.

To obtain a measure of the angle between two lines in a plane (and we are here limiting ourselves to the case of plane curves) draw a circle (Fig. 24), of any radius r, from the point of intersection of the lines as centre. The angle θ is measured by the ratio which the length, s, of the arc of the circle intercepted between the lines bears to the radius. It follows that the ratio of θ to s is equal to $1/r$,

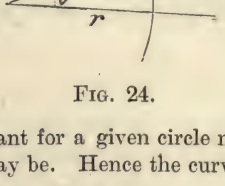


FIG. 24.

and is therefore constant for a given circle no matter how large or how small θ and s may be. Hence the curvature of a circle is the reciprocal of the radius.

Now it is always possible to draw a circle the curvature of which is the same as that of a given curve at a given point. This circle is called the *circle of curvature* at that point; its radius is called the *radius of curvature;* and the reciprocal of its radius measures the curvature of the given curve at that point.

In considering change of velocity as dependent on change of direction, it will be convenient to assume first that the speed is constant, and also that the rate of change of direction is constant. In other words, we shall investigate the case of uniform motion in a circle. Draw any two diameters at right angles to each other (Fig. 25), and let α, β be unit vectors along them. Let ρ be the vector of the point P. According to the data OP revolves uniformly, that is, the angle through which it turns per unit of time (called the *Angular Velocity*) is constant. Let ω be the angular velocity, so that, if P starts from the point A, the value of the angle POA is ωt, where t is the time

taken by the point to travel over the distance AP. Then we have vector ON = OP cos ωt . α, and vector NP = OP sin ωt . β, so that if OP = a

$$\rho = a(\cos \omega t \,.\, \alpha + \sin \omega t \,.\, \beta)$$
$$\dot{\rho} = a(-\omega \sin \omega t \,.\, \alpha + \omega \cos \omega t \,.\, \beta)$$
$$\ddot{\rho} = -\omega^2 a(\cos \omega t \,.\, \alpha + \sin \omega t \,.\, \beta)$$
$$= -\omega^2 \rho.$$

This result shows that the direction of the acceleration is *inwards* (for the negative sign is used) along OP, and that its magnitude is the square of the angular velocity multiplied by the tensor (§ 42) of ρ. But the tensor of ρ is a, and the angle described per unit of time is the speed, v, of P in the curve divided by a. Hence ω^2 is equal to

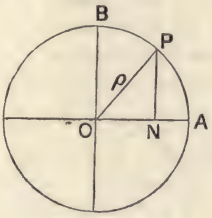


FIG. 25.

v^2/a^2, and therefore the magnitude of the acceleration is v^2/a. This acceleration, it is to be observed, does not alter the speed, but only the direction of motion; the reason being that it is perpendicular to that direction.

If OP be constant in direction, but of varying length r, the acceleration is $\ddot{r}$; and we have just found that, if r is fixed in length but revolves with angular velocity ω ($= \dot{\theta}$, if OP makes an angle θ with a fixed line), the acceleration is $-r\dot{\theta}^2$. Hence, if both magnitude and direction vary, the acceleration along r is

$$\ddot{r} - r\dot{\theta}^2.$$

44. *Acceleration in, and perpendicular to, the Direction of Motion.*—When a point moves in any curve, the acceleration perpendicular to the direction of motion at any position may be found by drawing the circle of curvature at the given position. If r is the radius of curvature, and v the speed of motion, the acceleration perpendicular to the direction of motion is, by last section, v^2/r towards the centre of curvature, and the acceleration of speed is $\ddot{s}$

where s is the distance measured along the curve, in the direction of motion, from a fixed point to the moving point.

The following method of deducing these results is exceedingly simple and instructive. Let ρ be the vector to any point of the path. Then

$$\dot{\rho} = \frac{d\rho}{dt} = \frac{d\rho}{ds} \cdot \frac{ds}{dt} = \frac{d\rho}{ds} \cdot v = v\rho' \text{ (say).}$$

Now, $d\rho$ being a vector, ρ' is also a vector in the same direction—that is, along the tangent (see Fig. 22, in which PP′ must be supposed to be indefinitely small). But, in the limit, when ds vanishes, the length of $d\rho$ is equal to ds. Therefore ρ' is a unit vector along the tangent. And, since the length of ρ' is constant, $d\rho'$ must be perpendicular to ρ'. Hence, $d\rho'/ds = \rho''$ is a vector inwards along the normal to the curve. And the quotient of the length of $d\rho'$ by the length of ρ' is equal to the angle, $d\theta$, turned through by the radius of curvature. Hence the magnitude of ρ'' is $d\theta/ds$; that is, (§ 43,) it is the reciprocal of the radius of curvature. But

$$\ddot{\rho} = \dot{v}\rho' + v\dot{\rho}' = v\rho' + v^2\rho''.$$

Hence, the total acceleration is made up of a part $\dot{v}$ along the tangent, and a part, proportional conjointly to v^2 and to the reciprocal of the radius of curvature, inwards along the normal.

45. *Average Speed and Velocity.*—If a point passes over a certain distance in a certain time with varying speed, it is always possible to find a uniform speed with which the same distance might have been described in the same time. This is called the *Average Speed* of the point. The last equation of § 41 obviously applies to the case of average speed.

When the acceleration of speed is uniform, the average speed is clearly half the sum of the initial and final speeds during the time considered. We may test this by the equations of section 42. V is the initial speed of projection, and $V - gt$ is the final speed. Hence the average speed is $V - \frac{1}{2}gt$. Therefore, by the last equation of § 41, the distance x is equal to $(V - \frac{1}{2}gt)t$, which agrees with the result already obtained.

The same results evidently hold for the corresponding angular quantities; for, in § 42, x might be an angular distance.

46. *Resolution and Composition of Velocities and Accelerations.*—Velocities and accelerations, since they are vector quantities (§ 40), are to be compounded and resolved in accordance with the laws of composition and resolution of these quantities. Hence, if a point is subjected to a series of simultaneous velocities which are represented by all the sides AB, BC, etc. (Fig. 26), of a closed polygon, taken in the

same order round, except one, the resultant velocity is represented by the remaining side taken in the opposite direction round. This theorem is known as the 'polygon of velocities.' It follows that, if the various velocities to which a point is subjected are representable by all the sides of a closed polygon taken in order, the point is at rest. For the resultant of all but one is equal and opposite to that one.

In the particular case of two velocities $\overline{AB}$ and $\overline{BC}$, the resultant is $\overline{AC}$—the third side taken in the opposite direction round. This

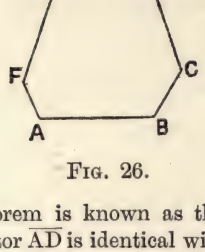


FIG. 26.

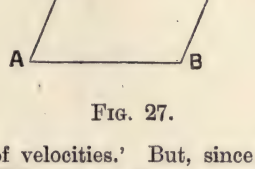


FIG. 27.

theorem is known as the 'triangle of velocities.' But, since the vector $\overline{AD}$ is identical with the vector $\overline{BC}$ (Fig. 27), we may say that the resultant of two velocities represented by adjacent sides of a parallelogram is the diagonal drawn from the same point. In this form the theorem is known as the 'parallelogram of velocities.'

In order to resolve a velocity into any number of components we have merely to reverse the above process. The problem is

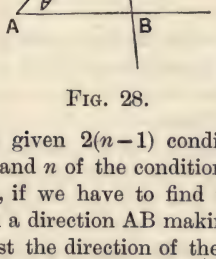


FIG. 28.

determinate if we are given $2(n-1)$ conditions, where n is the number of components and n of the conditions are directional.

As a particular case, if we have to find the resolved part of a velocity $\overline{AC}$ (Fig. 28) in a direction AB making an angle θ with AC, we require to know first the direction of the other component. It is usually understood that the other component is to be at right angles to the first, in which case we have $AB = AC \cos \theta$.

These remarks apply to accelerations and to any other vector quantities.

47. *Motion of Projectiles.*—Let α, β, be unit vectors in the horizontal and vertical directions respectively, and let a point be projected from O (Fig. 29) with velocity $a\alpha + b\beta$. If P be the position of the point at time t, ρ and ρ' being the components of the vector from O to P, we have

$$\rho = at\alpha$$
$$\rho' = bt\beta - \tfrac{1}{2}gt^2\beta.$$

In the latter equation, the first term represents the vector height to which the point would have ascended had gravity not acted; and

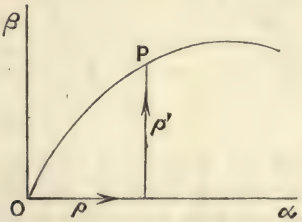


FIG. 29.

the second term represents the extent (§ 42) to which gravity has diminished this height. The length of ρ' is therefore $bt - \frac{1}{2}gt^2$, and this vanishes when $t = o$ and $t = \frac{2b}{g}$. The value $t = o$ corresponds to the instant of projection, and the other value gives the time of flight on a horizontal range. Again

$$\dot{\rho}' = b\beta - gt\beta.$$

This vanishes when $t = b/g$. But $\dot{\rho}'$ ceases to alter in magnitude just at the instant that the highest point of the path is reached. Hence, b/g is the time taken to reach the greatest height, and is equal to one-half of the time of flight.

If this value of t be put in the expression for ρ', we find that the total height reached is $b^2/2g$. Also, by putting $t = 2b/g$ in the expression for ρ, we find that the total range along a horizontal line is $2ab/g$.

48. *The Hodograph.*—If, from any point as origin, a line be drawn to represent the velocity of a moving point, the free extremity of the line traces out a curve which is called the *hodograph* of the

path of the moving point. The tangent to the hodograph gives the direction of acceleration in the path.

When the hodograph and the law of its description are given, the path and the law of its description can be found. Thus, in the path of a projectile, we have as above $\overline{OP} = \sigma = \rho + \rho' = ata + (bt - \frac{1}{2}gt^2)\beta$; and therefore (§ 41) the vector in the hodograph is $\dot{\sigma} = a\alpha + (b - gt)\beta$. It is consequently a vertical straight line uniformly described. And from this latter equation the former (which is its integral) can be obtained.

49. *Moments.*—The moment of any quantity is the measure of its importance with regard to the production of some effect. The

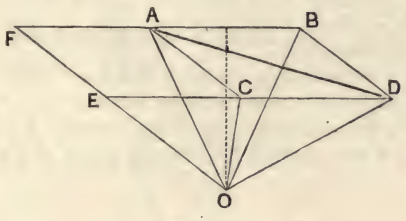


FIG. 30.

moment of any directed quantity (which may be indicated by the line AB) with reference to revolution about a point O is proportional to its own magnitude and to the distance of O from its line of action. If we define unit moment as the moment of a directed quantity of unit magnitude about a point at unit distance from its line of action, pa is the magnitude of the moment of a quantity containing a units about a point distant p units of length from its line of action. Thus the moment of AB about O is twice the area of the triangle AOB.

The moment of the resultant of two directed quantities is the sum of the moments of the components.—Let AC, AB (Fig. 31), be the two components; AD being the resultant. We have to prove that AOD = AOB + AOC. Draw OF parallel to CA to meet BA and DC produced in F and E. Then AOD = AOB + BOD − ABD = AOB + $\frac{1}{2}$ FEDB − $\frac{1}{2}$ ACDB = AOB + $\frac{1}{2}$ FECA = AOB + AOC.

The lines AC and AB in the figure have been so drawn that motion along them from A involves revolution in the same direction about O. Had AC been so drawn as to involve rotation about O opposite to that indicated by AB, it would have been necessary to regard one of the triangles as being *negative*. The same proof would then hold. It is usual to regard rotation opposite to that of the hands of a watch as positive.

When the direction of the quantity passes through O, its moment about O vanishes.

50. *Acceleration Perpendicular to Radius-vector.*—In § 43 we obtained an expression for the accleration along the radius-vector of a moving-point. We can now find an expression for the acceleration perpendicular to the radius-vector.

If AB, in the last figure, represents the path of a moving point P, the moment of the velocity of P is twice the rate at which the area AOP is described as P moves along AB. For if δs is a small length measured along the path, $p\delta s = p\frac{\delta s}{\delta t}\delta t$ is twice the corresponding increase of the area ($=\delta a$, say). Therefore $\frac{\delta a}{\delta t} = p\frac{\delta s}{\delta t}$, and, in the limit when $\delta t = o$, $\frac{da}{dt} = p\frac{ds}{dt} = pv$, where v is the speed, which proves the statement. It is evident that the proof still holds when the direction of motion varies, for it is true however small δs and δt may be.

Let OP($=r$), the radius-vector of the moving point P (Fig, 31), make an angle θ with a fixed line in the plane of the figure. Let

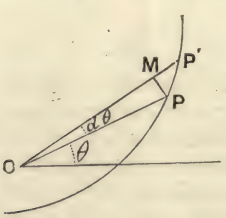


FIG. 31.

P move to a point P′, so near to P that PP′ is practically a straight line. Draw PM perpendicular to OP′. Then PM$=rd\theta$, and OPM$=\frac{1}{2}r\,.\,rd\theta$. Hence the area traced out per unit of time is $\frac{1}{2}r^2\dot{\theta}$, so that $r^2\dot{\theta}$ expresses the moment of the velocity. Now we have just seen that the moment of the velocity is equal to pv. But pv is equal to ru, where u is the resolved part of the velocity perpendicular to r. This is evident since, if p and r are inclined at an angle ϕ to each other, we have $p = r\cos\phi$, and $u = v\cos\phi$. Hence we have

$$ru = r^2\dot{\theta}.$$

Suppose now that the velocity is accelerated. The moment of the velocity represents the rate of increase of the area, and so the moment of the acceleration gives the rate of change of da/dt.

The acceleration can be regarded as composed of two parts—one part perpendicular to r, and the other along r. Of these components the former alone is effective in altering the rate of description of the area, for, § 49, the direction of the latter part passes through O. Hence the moment of the acceleration is $r\dot{u}$, and so the acceleration perpendicular to the radius-vector is

$$\dot{u} = \frac{1}{r}\frac{d}{dt}\cdot r^2\dot{\theta}.$$

51. *Simple Harmonic Motion.*—When a point P revolves with uniform speed in a circle, the motion of the foot N (Fig. 32) of the perpendicular drawn from P to any fixed diameter is called simple harmonic motion. The velocity and acceleration of the point N can easily be found when the position and velocity of P is given.

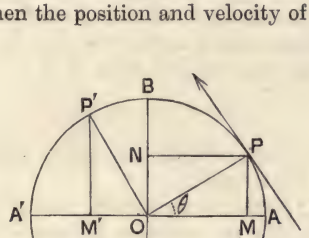


FIG. 32.

The velocity of N is evidently the resolved part of the velocity of P in the direction ON. That is to say, its magnitude is $v\cos\theta$ if v is the speed of P. But $\cos\theta$ is proportional to NP; hence the speed of N is proportional to NP. The maximum speed is attained when N passes through O; and it is then equal to v, the speed in the circle.

Similarly the acceleration of N is the resolved part of the acceleration of P in the line ON. But the acceleration of P is, § 43, $-v^2/a$, where a is the radius. Hence the acceleration of N is $-v^2/a \cdot \sin\theta = -v^2\,\mathrm{ON}/a^2$. That is to say, it is inwards towards O, the centre of the range of N; and its magnitude is proportional to ON, the displacement from the centre.

The ratio of the acceleration to the displacement is $v^2/a^2 = \omega^2$, where ω is the angular velocity of OP. But the angular velocity

is $2\pi/\tau$, τ being the time of a complete revolution in the circle. Hence

$$\frac{\text{acceleration}}{\text{displacement}} = \frac{4\pi^2}{\tau^2}.$$

If we call B the positive extremity of the range, the angle through which OP has turned since it coincided with OB is called the *Phase* of the simple harmonic motion. The phase may also be measured in fractions of a whole revolution. Thus we speak of the quarter phase, etc.

The maximum distance to which N can get from O is called the *Amplitude* of the motion. It is obviously the radius of the corresponding circle.

If the motion does not commence at the positive extremity of the range, the angle through which the radius has to turn until P reaches the positive end is called the *Epoch*. Thus the general expression for x, the distance of N from O, is

$$x = a \cos(\omega t + \alpha).$$

Here a is the amplitude, ω is the (constant) angular velocity, t is the time, and α is the epoch.

Simple harmonic motion is frequently exemplified in nature. It occurs in the vibration of stretched strings, the agitation of the luminiferous medium, etc.

52. *Composition of Simple Harmonic Motions.*—(1) *Motions in the same straight line and of equal periods.* Let the motion of P (Fig. 33) correspond to one of the given motions. From P draw a line PQ, making an angle ϕ, equal to the phase of another of the motions, with the line OA; and let the length of PQ be equal to the amplitude of this motion. Since ϕ and θ increase at the same rate, the line OQ remains of constant length and revolves at the same rate as OP and PQ. But the foot of the perpendicular from Q on OB moves with a motion which is compounded of the two given motions. Hence the resultant of the two motions is another simple harmonic motion, of the same period, in the same straight line. And, in particular, when the amplitudes of the two components are equal, the phase of the resultant is the mean of the phases of the components.

This proof is quite general, and applies to any number of such simple harmonic motions.

Another proof may be obtained as follows. Let the separate motions be $x_1 = a_1 \cos(\omega t + \alpha_1)$, $x_2 = a_2 \cos(\omega t + \alpha_2)$, etc. Then $x = x_1 + x_2 +$ etc. $= a_1 \cos(\omega t + \alpha_1) + a_2 \cos(\omega t + \alpha_2) +$ etc. But

$\cos(\omega t + \alpha_1) = \cos \omega t \cos \alpha_1 - \sin \omega t \sin \alpha_1$. Therefore $x = (a_1 \cos \alpha_1 + a_2 \cos \alpha_2 + \text{etc.}) \cos \omega t - (a_1 \sin \alpha_1 + a_2 \sin \alpha_2 + \text{etc.}) \sin \omega t$. Now assume the multiplier of $\cos \omega t$ to be equal to $a \cos \alpha$, and the multiplier of $\sin \omega t$ to be $a \sin \alpha$, so that $x = a \cos \omega t \cos \alpha - a \sin \omega t \sin \alpha$. This gives $x = a \cos(\omega t + \alpha)$. [The assumption made is obviously justifiable; for it only introduces two new quantities a and α, and gives two equations to determine their values.] Hence, in this case, the resultant is simple harmonic motion of the same period in the same straight line.

(2) *Two simple harmonic motions, of the same period and phase, in lines inclined at any angle.* These obviously compound into a single motion of the same period and phase. For, let OA, OB be the two inclined lines, and draw any other line OP. From P draw PM, PN, parallel to OA, OB, respectively, to meet these lines in the points N and M respectively. If P moves along OP with simple

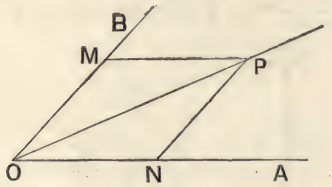


FIG. 33.

harmonic motion, it is evident that M and N move along OB and OA with simple harmonic motions of the same period and phase. And the motion of P is the resultant of their motions.

(3) *Two simple harmonic motions in lines at right angles to each other of the same period and amplitude, and differing in phase by one quarter of a period.* A glance at the figure of last section shows that M′ moves along OA′ precisely as N moves along OB, since OP′ is perpendicular to OP. Hence the motions of N and M differ in phase by one quarter of a period. And the motion of P is their resultant. That is to say, the resultant is uniform circular motion; and the motion takes place from the positive end of the range in which the motion is one quarter of a period in advance to the positive end of the other range.

(4) *Two simple harmonic motions of equal period, and of phases differing by a quarter period, in lines inclined at any angle.* By projection of the circle we obtain an ellipse in which conjugate diameters correspond to mutually perpendicular diameters of the circle. Hence the resultant is elliptic motion, with equal areas

described in equal times (since this is so in the circle), and with a law similar to that given in (3) as regards direction of motion.

(5) *Any number of simple harmonic motions, in lines inclined at any angle to each other, and of any phase, but of equal periods.* By a reversal of the first proof of (1), we see that the line OQ may be replaced by *any* two lines OP and PQ, which revolve with the same angular velocity. Hence any simple harmonic motion may be broken up into two of the same period, which differ in phase by any given amount, and one of which has any given phase. This appears also by a reversal of the second proof of (1). For if $a \cos(\omega t + \alpha)$ is to be identical with $a_1 \cos(\omega t + \alpha_1) + a_2 \cos(\omega t + \alpha_2)$, we must have $a_1 \cos \alpha_1 + a_2 \cos \alpha_2 = a \cos \alpha$, and $a_1 \sin \alpha_1 + a_2 \sin \alpha_2 = a \sin \alpha$. That is, there are only two conditions to be satisfied by the four quantities a_1, a_2, α_1, α_2; so that two more may be imposed.

Let P_1, P_2, etc. (Fig. 34), be the points moving with simple harmonic motion. Let p_1, p'_1, be two points, whose motions com-

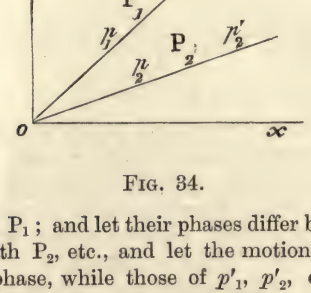


FIG. 34.

pound into that of P_1; and let their phases differ by a quarter period. Deal similarly with P_2, etc., and let the motions of p_1, p_2, etc., be all of the same phase, while those of p'_1, p'_2, etc., also agree in phase with each other, but, of course, differ in phase from the motions of p_1, p_2, etc., by a quarter of a period. Resolve all these motions into their components along two rectangular axes ox and oy. Then compound all of the same phase in each axis with each other. The result is, by (1), two simple harmonic motions in each axis differing in phase by a quarter period. Now combine each motion in oy with the motion in ox which is of the same phase, and we have ultimately two simple harmonic motions which differ in phase by a quarter period and take place in lines which are in general inclined to each other at a finite angle. These, as we have seen, combine into elliptic motion.

53. *Wave Motion along a Line.*—Let a point vibrate up and down the y-axis with simple harmonic motion about the origin, and let the paper be drawn along behind it at a uniform rate in the direction xo. The point will trace out a curve (indicated in the figure) which exhibits the simplest form of a wave. The quantities y and x can easily be seen to be connected by the relation

$$y = a \cos (\omega t - nx)$$

in which a and ω denote the same quantities as formerly. If x is constant, the equation is of the same form as the one which was given in § 51, and shows that every point on the x-axis vibrates up

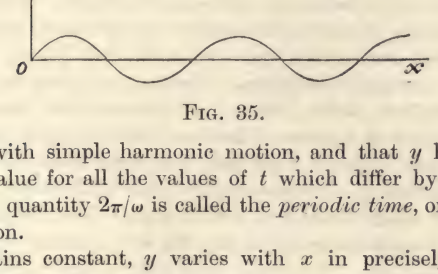


FIG. 35.

and down with simple harmonic motion, and that y has one and the same value for all the values of t which differ by the amount $2\pi/\omega$. This quantity $2\pi/\omega$ is called the *periodic time*, or the *period*, of the motion.

If t remains constant, y varies with x in precisely the same manner as it did when x was constant and t varied. The value of y is the same for all values of x which differ by $2\pi/n$. This quantity is the *Wave-length*.

Lastly, y remains fixed in value if x and t vary simultaneously in such a way that $(\omega t - nx)$ is zero, x being measured from any special position, and t from any definite instant. This gives $\dot{x} = \omega/n$, and shows that the wave runs along in the direction ox with speed ω/n and without change of form.

Similar reasoning shows that the equation

$$y = a \cos (\omega t + nx)$$

represents a succession of precisely similar and equal waves which run in towards the origin with speed $-\omega/n$. The resultant disturbance due to the superposition of this set of waves upon the former is represented by

$$\begin{aligned} y &= a \{\cos (\omega t - nx) + \cos (\omega t + nx)\} \\ &= 2a \cos \omega t \cos nx. \end{aligned}$$

Whatever be the value of x, this vanishes when t is any odd multiple

of $\pi/2\omega$; and, whatever be the value of t, it vanishes when x is any odd multiple of $\pi/2n$. The motion at any definite point is simple harmonic, of period $2\pi/\omega$; and the form of the wave at any definite instant resembles that shown in Fig. 35, the ordinates being all altered in the common ratio of 2 cos ωt to unity. The resultant is therefore a series of *stationary* waves, which oscillate up and down, parallel to the y-axis, *in situ*. This result is of importance in the theory of the vibrations of stretched strings, etc.

54. *Rotation.*—While a mere *point* can have translational motion only, a rigid body (a body the parts of which cannot suffer relative displacement) is free to *rotate* also unless three points of it, which do not lie in the same straight line, are fixed. Three such points being fixed, the body is devoid of all freedom to move. If two points are fixed it can rotate about the line which joins them, and is said to have one degree of rotational freedom. If one point only is fixed the body may rotate independently about any three mutually perpendicular axes which pass through that point—it has three degrees of rotational freedom. Finally, no point being fixed, it has, in addition, three degrees of translational freedom. The greatest number of degrees of freedom which a rigid body can have is therefore six. [A non-rigid body has distortional freedom also.]

55. *Alteration of Co-ordinates because of Rotation.*—If a point P is rotating about the axis of z (drawn perpendicular to the plane of the paper through the point O, Fig. 36) with angular velocity ω_z, the

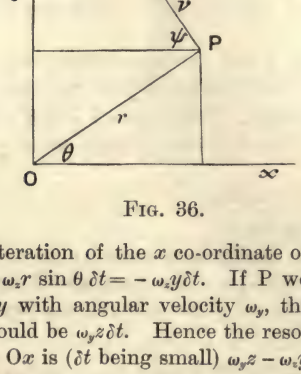


Fig. 36.

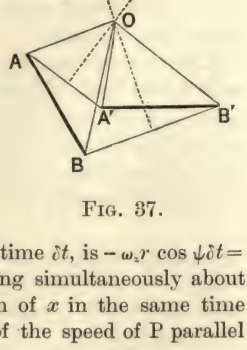


Fig. 37.

alteration of the x co-ordinate of P, in the time δt, is $-\omega_z r \cos\psi\delta t = -\omega_z r \sin\theta\,\delta t = -\omega_z y\delta t$. If P were revolving simultaneously about Oy with angular velocity ω_y, the alteration of x in the same time would be $\omega_y z\delta t$. Hence the resolved part of the speed of P parallel to Ox is (δt being small) $\omega_y z - \omega_z y$.

56. *Uniplanar Motion of a Rigid Body.*—By 'uniplanar motion' is meant motion parallel to one plane. The motion of a rigid plane figure is included as a special case.

Let the motion be parallel to the plane of the paper; and let AB be the position of a line in the body before the motion occurs, while A′B′ is its position at the end of the motion. Draw AA′ and BB′; bisect them, and erect perpendiculars at their points of bisection. Let these meet in O. We have OA = OA′, and OB = OB′. Also AOB and A′OB′ are congruent triangles, the angle AOB being equal to the angle A′OB′. Thus AB might have been moved into the position A′B′ by a single rotation, about O as centre, through the angle AOA′. Hence *any displacement of a rigid body parallel to one plane may be produced by rotation about a definite axis perpendicular to that plane.*

In general the body does not revolve in this way from its initial to its final position. On the contrary, each point usually describes a curve which is not the arc of a circle. In such a case we may regard the total displacement of any point as made up of a succession of indefinitely small displacements, each of which coincides with an indefinitely small arc of a circle. This circle is evidently the circle of curvature (§ 43) of the path of the moving point. Its centre is called the *instantaneous centre* about which all points in the same plane are revolving. Thus, when a wheel rolls along the ground, the point of the wheel which is in contact with the ground is at rest for an instant—it is the instantaneous centre about which the wheel is revolving for a moment as a rigid body.

Let the point O, Fig. 38, be the instantaneous centre. The point p_1 revolving about O will come into the position p'_1. Suppose now that p_2 revolves about p'_1 as the new instantaneous centre, and that this brings it into coincidence with p'_2 about which the revolution next

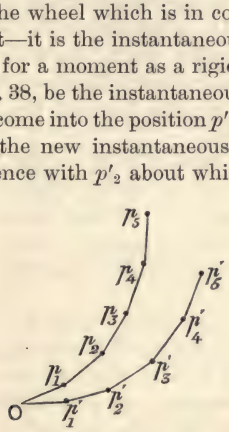


FIG. 38.

takes place, and so on. The points p_1, p_2, etc., are points, fixed in the body, which are successively at rest, and the points p'_1, p'_2, etc., are points, fixed in space, at each of which in succession the instantaneous centre is situated. The instantaneous centre coincides for a short time with each point of both series, and passes suddenly

from one to another. When the motion is continuous the polygons $op_1p_2 \ldots$ and $op'_1p'_2 \ldots$ become continuous curves, and the instantaneous centre moves continuously along them. In the case of a rigid body, moving parallel to the plane of the paper, the line through the centre perpendicular to that plane is instantaneously at rest, and is called the *instantaneous axis;* and the curves in the figure are sections of cylindrical surfaces in the body and in space. Hence we see that *the most general uniplanar motion of a rigid body consists in the rolling of a cylinder fixed in the body upon a cylinder fixed in space.* An obvious example of this is given when a roller is drawn over the surface of the ground.

A similar statement, modified merely by the substitution of the word 'curve' for 'cylinder,' applies to the motion of a plane figure in its own plane.

Mere *translation* is a special case in which the instantaneous centre is at an infinite distance. It may be considered to consist in infinitely slow rotation about an infinitely distant axis.

57. *Motion of a Rigid Body in Space.*—First, suppose one point of the body to be fixed, and consider a sphere in the body with its centre at the fixed point. Take any two points A,B, on the surface of this sphere which occupy the positions A',B', respectively, at the end of the motion. The reasoning of last section applies here also, great circles of the sphere taking the place of straight lines. We thus find that the displacement might have been produced by simple rotation of the sphere about a diameter passing through the point O (Fig. 37) on the surface.

The actual motion consists in the rolling of a cone fixed in the body upon a cone fixed in space. This is at once evident if we suppose the curve $op_1p_2 \ldots$ of last section to be drawn upon the surface of the sphere whose centre is fixed—the points p_1, p_2, etc., being successive positions of the extremity of the instantaneously fixed diameter.

Now suppose that no point is fixed. The total displacement consists in general of both translational displacement and rotational displacement; and it is clear that we may separate these, taking first one and then the other, and yet produce the same total effect. As we have just seen, the rotation leaves a set of planes (those perpendicular to the axis of rotation) parallel to their original position, and mere translation does not alter this parallelism; so that, in the final position of the body, there is one set of planes which has been unaltered as regards orientation. The required rotational displacement can be produced by revolution about *any* axis perpendicular to these unaltered planes, but the final *position* of the body

will depend on the particular axis chosen. And we can so choose the axis that, after the rotation has taken place, mere translation parallel to that axis will make the body take the required position. And this can only be done in one way. For let the plane of the paper be one of the planes which are unaltered in direction, and let AO be a line in that plane, the final position of which is to be A′O′ in a parallel plane. Rotation through the angle between AO and A′O′,

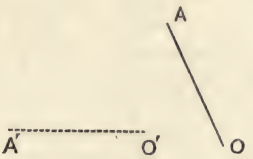


FIG. 39.

about *any* point in the plane of the paper, will make AO *parallel* to A′O′; but AO will only be superposed upon A′O′ if the point be chosen by means of the construction of last section. And, when this superposition is effected, a translation perpendicular to the plane of the paper will bring the body into its final position. *Hence any displacement of a rigid body in space may be produced by rotation about, and translation along, a definite axis.*

When the rotation and the translation are simultaneous, the motion is called a *twist about a screw*. Such a motion is the most general kind of motion that a body which possesses only one degree of freedom can have. [The rotation and the translation do not occur *independently*, and so one degree of freedom alone is involved.]

Any given motion of a rigid body in space consists in a twist about a screw, in which the axis and the linear and angular velocities are in general varying. The position of the axis at any instant is given by the line of contact of two ruled surfaces, one of which (fixed in the body) simultaneously rolls and slides upon the other, which is fixed in space.

58. *Composition of Angular Velocities.*—An angular velocity is completely determined when its magnitude and direction are given, and these quantities may be indicated by the magnitude and direction of a line drawn parallel to the axis of rotation. Angular velocities (and accelerations) are therefore vector quantities, and their laws of composition and resolution are identical with the laws for vectors, and therefore with the laws for linear velocities and

accelerations. Thus the resultant of two angular velocities or accelerations, which are represented by the two sides of a parallelogram, is represented, on the same scale, by the conterminous diagonal.

An extremely important case is that in which a body with one point fixed is revolving uniformly about an axis and is subjected to uniform angular acceleration about a perpendicular axis. In the corresponding problem regarding linear velocity and acceleration (§ 43), the magnitude of the linear velocity is unaltered while the direction of motion revolves uniformly, being always perpendicular to the direction of acceleration, which also revolves uniformly. So, in the present case, we can at once assert that the magnitude of the angular velocity will remain constant, but that its axis will revolve uniformly, and will be always perpendicular to the axis of constant acceleration. Now the axis of acceleration is always horizontal when the rotating body is a top which spins uniformly with its axis of revolution inclined to the vertical. Hence the direction of the axis of the top will rotate uniformly, and will be always perpendicular to a horizontal line (*not* a horizontal plane) which rotates uniformly. The axis must therefore revolve at a uniform rate around the vertical.

Precession of the equinoxes is due to angular acceleration of the earth about an equatorial axis, and the peculiar motions of gyrostats have a similar explanation.

To compound two angular velocities about parallel axes, indicated in magnitude and direction by AB and CD (Fig. 40), it is merely necessary to find a point O such that the moments of AB and CD about it are equal and opposite (§ 49). Let p_1, p_2, be the lengths of perpendiculars from O upon AB and CD respectively. Since the angular velocity of O due to rotation about AB is numerically equal

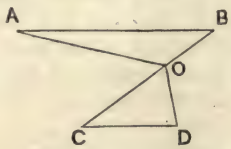


FIG. 40.

to AB, it follows that $AB p_1$ represents the velocity of O perpendicular to the plane of the paper; and, similarly, the velocity of O perpendicular to that plane due to rotation about CD is $CD p_2$; and this quantity is of the opposite sign to the former, for points moving

along AB and CD revolve oppositely around O. Hence, if the areas AOB and COD are equal, the point O is at rest; and thus the locus of O—that is, a line through O parallel to AB and CD—is the resultant axis of rotation.

Similar reasoning shows that the resultant axis due to angular velocities indicated by AB and DC is a parallel line situated above AB, and that the direction of rotation coincides with that of AB, the larger of the two components.

To find the effect of the superposition of an angular velocity ω upon a linear velocity v, or of a linear velocity v upon an angular velocity ω, we may resolve the linear velocity into its two components parallel to and perpendicular to the axis of rotation. The effect of the perpendicular component v' is to shift the axis parallel to itself through a distance d, such that the speed $d\omega$ is equal and opposite to the speed v'. The parallel component simply moves the whole body in the direction to the axis, so that the resultant is a twist.

If a body is rotating simultaneously about three axes, the velocities being representable by the three sides of a triangle taken in the same direction round, the effect is that there is no rotation; but the body is translated perpendicularly to the plane of the triangle with a speed represented by twice the area of the triangle. If one of the rotations is represented by a side of the triangle taken in the opposite direction round, there is no translation; but the body rotates about an axis bisecting the two sides which were taken in the same way round. The rotation round this axis coincides in direction with the rotation about the side of the triangle parallel to it, and the angular velocity is twice as great as the velocity about the parallel side.

Corresponding results obtain in the case of angular velocities representable by the sides of closed plane polygons.

59. *Displacement of the Parts of a Non-Rigid Body.—Strain.*—A non-rigid body may alter in form, or in volume, or both. Any such definite change of shape or bulk is called a *Strain*.

Homogeneous Strain.—When all parts of a body, originally similar and equal, are similarly and equally strained, the strain is said to be *homogeneous*. It follows that parallel straight lines in the unstrained body become parallel straight lines in the strained body; but, in general, the direction of the lines and the distance between them is altered. And therefore parallelograms remain parallelograms, parallelepipeds remain parallelepipeds, and any figure or surface changes into a similar figure or surface. Thus a sphere becomes an ellipsoid.

Such a strain is completely determined when we know what

alteration—in magnitude and direction—has been produced in three originally non-coplanar lines. One number is required for each line to fix the change of length; and two numbers are required to determine the change of direction of each line. In all, nine numbers are in general necessary.

The simplest kind of homogeneous strain is that in which there is uniform expansion or compression in all directions. Any line in the strained figure preserves its original direction, and all lines are equally altered in length. Such strain occurs in the compression of fluids. One number completely determines it.

Next in order of simplicity is a homogeneous strain in which lines in one definite direction are unaffected by the strain; that is, the strain is confined to planes perpendicular to a definite direction, and the alteration in any one of these planes is precisely similar and equal to the alteration in any other. It is usual, therefore, to call such a distortion a *plane strain*. A circle in the unstrained figure, drawn in one of the planes of distortion, becomes an ellipse; and so a sphere in the body becomes an ellipsoid. (In the cases of equal expansion or of equal contraction in all directions in the planes of the strain, the ellipsoid is an oblate or a prolate spheroid respectively. All lines which are not in or perpendicular to these planes are altered in direction.) All perpendicular diameters of the circle become conjugate diameters of the ellipse; in particular, the principal axes of the ellipse were originally perpendicular diameters of the circle—from which, however, they usually differ in direction. Four numbers determine a plane strain of the most general kind.

When the principal axes of the ellipse (called the *strain-ellipse*) into which the circle is deformed are not changed from their original directions in the unstrained body, the strain is called a *pure* (*or non-rotational*) *plane strain*. In this case the distortion consists in extension (or contraction) in two directions at right angles to each other. Any rotational, or impure, plane strain may (so far as the final effect is concerned) be produced by a pure strain superposed upon, or followed by, rotation about a definite axis perpendicular to the plane. In a pure strain every line except a principal axis has suffered rotation; and it follows from this that the superposition of two pure strains generally produces an impure strain. Hence a body may be distorted by three plane strains in succession, and yet (the strains being properly chosen) be left unstrained, but rotated through an angle about a definite axis. Three numbers completely characterise a pure plane strain.

A specially important case is that in which there is no alteration

of volume. This implies elongation in one direction and equal contraction in another. We may suppose that these directions are mutually perpendicular, giving a pure strain; for, as we have seen, any other strain may be assumed to consist in a pure strain followed by rotation as of a rigid body. Let oy be the direction of the elongation, and let ox be the direction of contraction. Let $abcd$ be

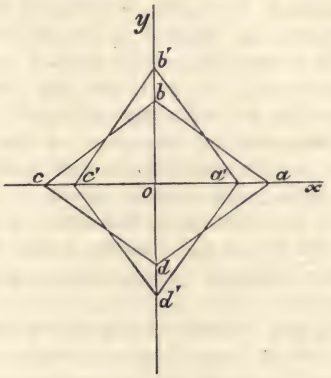


FIG. 41.

a rhombus in the unstrained figure, which becomes the rhombus $a'b'c'd'$ in the strained state, oa being equal to ob', oa' being equal to ob, and so on. We have then ab equal to $a'b'$, with similar results for the other sides of the rhombus. It is, therefore, obvious that there are two sets of planes in the figure which experience no alteration, except as regards position; so that (rotation excluded) the strain

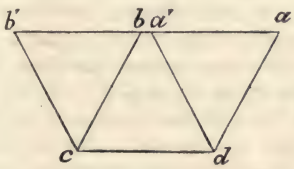


FIG. 42.

might have been produced by holding fast one plane of either set—say the plane through cd perpendicular to the plane of the paper—and sliding all planes parallel to it through a distance proportional to their distance from the fixed one, until the originally acute angle

of the rhombus becomes equal to its supplement. This motion is termed shearing motion, and the strain is called a *simple shear*.

To make the result of this shear coincide with the result of the above pure strain, we must turn the body round in the direction of the hands of a watch through an angle $b'db$, so that $b'd$ coincides with bd—that is, through an angle equal to half the difference between the obtuse and acute angles of the rhombus.

In the most general homogeneous strain, a sphere in the unstrained body becomes an ellipsoid in the strained state. Any set of three mutually perpendicular axes become mutually conjugate diameters of the ellipsoid. In particular, the three principal axes of the ellipsoid (which are called the principal axes of the strain) were originally perpendicular diameters of the sphere. These principal axes are usually rotated from their initial positions; and, as in the corresponding case of plane strain, when this turning of the principal axes does not occur, the strain is said to be *pure* or *non-rotational*. So far as the ultimate result is concerned, any impure strain may be looked upon as due to a pure strain followed by a rotation.

Any given strain may be produced by a simple shear followed by an extension (or contraction) perpendicular to the plane of the shear which in turn is succeeded by a uniform expansion (or compression). For the shear may be continued to such an extent as to give the proper ratio of the maximum and minimum axes; and the perpendicular extension will then give the proper ratio of the mean axis to each of the other two; while, lastly, the uniform expansion can be continued to such an extent as to give the proper magnitudes of the axes.

60. *Non-Homogeneous Strain.*—So long as rupture does not occur, all displacements in a portion of matter are essentially continuous. Hence, however greatly the displacements may change throughout the body, we can always consider a portion so small that, within its limits, the strain is homogeneous.

Let P and Q be two near points—so near that the strain is homogeneous, and let δx, δy, δz, or ξ, η, ζ, be the co-ordinates of Q relative to P. If u, v, w, be the components of the displacement of P parallel to the axes of x, y, and z, respectively, we may denote by $u+du$, $v+dv$, $\omega+dw$ the corresponding quantities for Q, so that du, dv, dw are the components of the relative displacement of P and Q. Each of these components will in general depend upon the values of ξ, η, and ζ; and each must be a *linear* function (§ 29) of these quantities since, in homogeneous strain, straight lines remain straight lines. Now, du/dx being the x-rate of variation of u, $du/dx \,.\, \delta x$ is the change of u due to the change δx, and so on;

so that $du = du/dx \,.\, \delta x + du/dy \,.\, \delta y + dw/dz \,.\, \delta z$. But, instead of du, we may write the equivalent quantity $d\xi$, and so finally

$$d\xi = \xi \frac{du}{dx} + \eta \frac{du}{dy} + \zeta \frac{du}{dz},$$

$$d\eta = \xi \frac{dv}{dx} + \eta \frac{dv}{dy} + \zeta \frac{dv}{dz},$$

$$d\zeta = \xi \frac{dw}{dx} + \eta \frac{dw}{dy} + \zeta \frac{dw}{dz}.$$

The quantities u, v, and w being known functions of x, y, and z, these equations enable us to determine fully the nature of the strain in the neighbourhood of any given point.

The multipliers of ξ, η, and ζ, in these equations, are the nine numbers which determine the strain (§ 59).

61. *Motion of Fluids.*—While the parts of a rigid body cannot suffer relative displacement, the parts of a non-rigid solid can be displaced relatively; but the magnitude of any displacement is not unlimited, for, if it be too large, the body will be ruptured. In an infinite expanse of fluid there is no limit to the possible increase of distance between two originally near parts.

A *Line of Flow* in a moving fluid is defined as a line so drawn that its direction at any point coincides with the direction of motion of the fluid at that point. It may, or it may not, be the actual path of any particle of the fluid. In illustration of this we may consider the motion of points in a spinning top (§ 58). At any instant the line of flow of a given point is a circle drawn round the axis of the top. But the axis is itself in motion, so that the path of the point merely coincides with the line of flow for an indefinitely small distance. Similarly, when the lines of flow in a fluid are in motion, the path of any particle only coincides with one line of flow for an indefinitely small interval of time.

When the lines of flow are fixed, so that they are actual paths of particles, the motion is said to be *steady*, and the lines are called *stream-lines*.

If lines of flow are drawn through all points of a closed curve, a *Tube of Flow* is formed. None of the fluid inside such a tube ever passes out of it, and none ever enters it from the outside.

Whatever be the nature of the strain throughout a solid, we may consider, instead of the total strain, the strain produced in a given indefinitely small period of time. The displacements in that period are evidently proportional to the instantaneous velocities

of the various parts. And, hence, all the results which we have obtained regarding displacements in a non-rigid solid have a direct application in the discussion of fluid motion.

Just as rotational strain may exist in a solid, so, in a fluid, there may be rotational, or, as it is termed, *Vortex Motion*. A line, drawn in the fluid so that its direction coincides at any point with the direction of the axis of rotation at that point, is called a *vortex line*. And a tube formed by vortex lines drawn through all points of an infinitely small closed curve forms a vortex tube, and is said to enclose a *vortex filament*. In most cases of fluid motion in which vortices exist, the vortically moving parts occupy only a small proportion of the whole volume of the fluid.

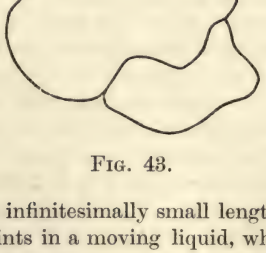


FIG. 43.

If ds represents an infinitesimally small length of a curve drawn between any two points in a moving liquid, while v represents the velocity parallel to the curve at any point, the integral of the quantity vds is called the *Circulation* along the curve. If we surround the curve by a tube (Fig. 43), the sectional area of which is numerically equal to the speed along the curve, the volume of that tube is numerically equal to the circulation along the curve. The positive or negative sign must be attached according as the circulation is in the positive or negative direction round the curve. When the curve is closed and lies in a plane, we may speak of the circulation 'round the enclosed area.' It is evident that the circulation round any area is equal to the sum of the circulations round its parts, for the circulations round a common boundary are equal and of opposite sign.

Shearing Motion.—The results which we have obtained regarding pure homogeneous plane strains apply directly to this case of fluid motion. And it is easy, in addition, to deduce useful results regarding the circulation of the fluid.

(1) *The circulation along any two plane conterminous curves, which nowhere lie at a finite distance from each other, is the same.*

To see this, let *abc* and *ac* be portions of two such curves, and let the straight line *mn* indicate the velocity v, which is constant all along these portions provided they are small enough. Hence the circulations round *abc* and *ac* are the products of v into the projections of *abc* and *ac* respectively upon *mn*. But these projections are equal, which proves the result.

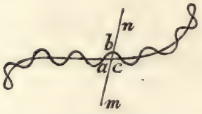


FIG. 44.

FIG. 45.

(2) *The circulations round any two similarly situated and equal plane areas are equal.* For the velocities in the one curve relatively to the point o are equal to those in the other curve relatively to the corresponding point o'. Hence the circulation round s' differs from the circulation round s by the product of the relative speed of o' and o into the projection of s' upon the line of relative motion of o and o'. But this vanishes, since s' is a closed curve.

(3) *The circulation round any plane curve is proportional to its area.* For we may divide the area into a series of indefinitely small, similarly situated, and equal, parallelograms. The circulation

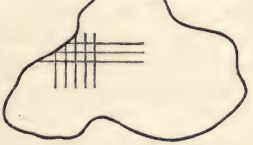



FIG. 46.

round each of these is equal by (2). And now, since the edge of the area formed by these parallelograms is nowhere finitely apart from the given curve, the result follows from (1).

(4) *The circulation round any plane area is equal to twice the area multiplied by the angular velocity of the fluid round a perpendicular axis.* Let the area be an indefinitely small circular area. From the analogy to strain, we see that any instantaneous motion may be broken up into a pure part and a rotational part. The

motion corresponding to the pure part is constant over the small area, and therefore contributes nothing to the circulation. The rotational part gives a tangential speed ωr, where ω is the angular velocity round a perpendicular axis through the centre, and r is the radius of the circle. Hence the circulation is $2\pi r \,.\, r\omega = 2\pi r^2\omega$. But πr^2 is the area of the circle, and so the proposition is true in this case; and, ω being constant, we see, by (3), that it holds in all cases.

Heterogeneous Motion.—A portion of the fluid may be taken so small that the motion is homogeneous throughout. The previous results are then applicable.

Reasoning similar to the above shows that the circulation round any closed curve (plane or not) in the fluid is equal to twice the integral of the normal angular velocity over any surface bounded by the curve. For a sufficiently small portion of the surface may be assumed to be plane, and the result (4) applies. Hence, if a closed surface be taken, and the normals at all points be drawn outwards alone, or inwards alone, the integral of the angular velocity over the surface is zero. For if we draw any closed curve round the closed surface, the integral over each portion of the surface corresponds to equal but opposite circulation round the closed curve.

Vortex Motion.—The above conclusion may be applied to the case of a finite portion of a vortex-tube. The sides of such a tube are parallel to the axes of rotation. Therefore the ends only contribute to the integral of the angular velocity; and so the integral for each end must be equal in magnitude, and it will be of the same sign in both cases if the normals are drawn in the same direction along the tube. Hence the circulation is the same at all sections of a vortex-tube. The tube being small, the angular velocity is therefore inversely proportional to the cross-section. Hence the vortex rotates faster the thinner it is.

We conclude also that a vortex must either return into itself, forming a closed circuit, or that its ends must be at the surface of the liquid; for the velocity of rotation can neither abruptly change nor become infinite within the liquid. A smoke-ring exemplifies the former case; the eddies formed round the edge of the hand, when it is dipped into water and drawn rapidly along, illustrate the latter.

CHAPTER VI.

MATTER IN MOTION.

62. *Force.*—The fundamental property of matter, which distinguishes it from the only other real thing in the universe, is *inertia*. And, in consequence of inertia, when we move a body we are conscious of making some exertion, and are accustomed to say that we exert *force*. Hence it is usual, as Newton did, to speak of force as the cause of motion; and the force may be of the nature of a push, a pull, an attraction, a repulsion, etc.

We do not yet know the nature of the physical process going on in matter which is in a state of tension, and so we figure it to ourselves by means of the mental impression caused by the muscular sense. But loudness and brightness, though they, as we shall see later, are mere subjective impressions, yet correspond to certain physical realities. We may therefore proceed to inquire whether or not there is some physical process corresponding to the impression of force.

We have already obtained (§ 7) a kinetic measure of energy; and by means of the new idea of force, we can now deduce a statical measure of it. Work is done when we move a body against the action of a force which we assert to be the cause of motion in the opposite direction. If the force is constant, we know that the work which is done is proportional to the distance through which the body is moved against the action of the force, for every equal addition to the distance is made under precisely similar circumstances. Also we know that the amount of work which is done is proportional to the force, it being more and more difficult to produce the displacement according as the opposing force is greater. Hence, provided we define the unit of work as the work done by unit force acting through unit distance, we may write

$$w = fs,$$

where w, f, and s represent respectively the work, the force, and

the distance. But, by the principle of conservation of energy, if a mass m is projected with speed v, and is brought to rest by the action of the force, we know that the work is equivalent to the kinetic energy which is lost. Therefore

$$\tfrac{1}{2}mv^2 = fs.$$

From this equation, denoting the energy by e, we deduce

$$de/ds = f.$$

That is, force is the space-rate of variation of energy.

Force is not conserved as energy is, although it may possess that kind of conservation spoken of in § 12. Indeed, Newton's third law of motion asserts that the total algebraic sum of the forces in the universe is zero.

And now, having arrived at a clear understanding of what 'force' really is, we may use the word, or any of its special equivalents, in subsequent sections, and speak of force as the cause of motion, without producing confusion of ideas.

There is no such thing in nature as a material *point*. However small a particle of matter may be, it always has a finite surface, and occupies a certain volume. And no actual force acts at a point merely; it is distributed throughout a volume, or is applied over a surface. As an example of the former class of forces, we may take the force of gravitation; as an example of the latter class, we may take the force of friction, *i.e.*, the tangential force which resists the sliding of portions of matter over each other. [This tangential force is independent of the area of the surface of contact of the two bodies—so long, at least, as the surface of contact is not so small that sliding motion cannot occur without producing abrasion of the substances. It is in general proportional to the normal pressure between the bodies; but, in many substances, it depends greatly upon the time during which the contact has lasted. It may be much reduced by the use of proper lubricants.

The law of friction may be expressed by the equation

$$\mathrm{F} = \mu' \mathrm{R},$$

where F is the force of friction, R is the normal pressure, and μ' is a constant, called the co-efficient of kinetic friction.

When the forces which act so as to produce motion are just insufficient to overcome friction, this equation becomes

$$\mathrm{F} = \mu \mathrm{R},$$

where μ—called the 'co-efficient of statical friction'—is, as experiment shows, a constant of greater numerical magnitude than μ'.

It follows that sliding motion will continue under the action of forces which were inadequate to start the motion.]

63. *The Laws of Motion.* — Newton's three Laws of Motion (expressed in terms of force regarded as the cause of motion) form at present the simplest foundation for the study of the phenomena of moving matter. The science which deals with these phenomena has, therefore, been called *Dynamics*, *i.e.*, the science which treats of the action of force upon matter. It is usual, also, to divide the subject into two parts—*Kinetics* and *Statics*—according as motion is, or is not, produced. It is impossible to doubt that ultimately a more fundamental, and at least equally simple, basis will be obtained in connection with the principles of energy. We can even at present make such a substitution for Newton's Laws, barring the simplicity.

64. *The First and Second Laws.*—The First Law asserts that *every body maintains its state of rest, or of uniform motion in a straight line, except in so far as it is caused by force to alter that state.*

The 'rest' here referred to is, of course, *relative* rest. And, instead of the last clause, we might say 'so long as it remains in a region of constant potential.' This law, in fact, asserts the conservation of the energy of the particular body considered, potential energy being regarded (§ 8) as energy which has passed to a connected system.

Uniformity of motion in magnitude and direction does not enable us to say that no force is acting upon the body, but only that the resultant of all the forces is zero—that they can be combined into two equal and opposite forces. This means that the body may be simultaneously gaining and losing energy at precisely equal rates.

The law implies, also, that force must be acting upon a body if the direction of its motion alters, the magnitude remaining constant. [In this case the body moves along an equipotential surface.]

The Second Law gives the relation between force and the effect which it produces. It states that *change of motion is proportional to force, and is in the direction in which the force acts.* By *motion* Newton meant what is now called *momentum*—the product of the mass and the velocity of the moving body. Of course, the change of momentum is proportional to the time during which a constant force has acted, so that we may express the second law by the equation

$$ft = mv,$$

v being the change of speed produced during time t in the given mass m by the average force f. The actual force at any instant is got by making the time indefinitely small, when the equation becomes

$$f = ma,$$

a being the acceleration. And this form of the equation may be used in all cases, whatever be the value of t, provided that f and a represent the average values of the force and the acceleration during the time t. Both equations involve the assumption (or, rather, the definition) that unit force is the force which, acting upon unit mass for unit time, produces unit change of speed. [From the latter we get $fds = mavdt = m\dot{v}vdt = mvdv$, and therefore $fs = \frac{1}{2}mv^2$, so that our new definition of force is equivalent to the former one (§ 62).]

If two forces, f_1 and f_2, act upon equal masses, the accelerations produced are proportional to those forces, for $f_1/f_2 = a_1/a_2$. Thus the Second Law of motion gives us a method of comparing forces. It also enables us to compare masses. For, if two equal forces act upon different masses, m_1 and m_2, we have $m_1a_1 = m_2a_2$; *i.e.*, the accelerations produced are inversely proportional to the masses.

By means of the first two laws alone we can investigate the motion of a material point or of a set of disconnected particles.

For example, in the equations of motion given in § 42, we have only to introduce the mass m as a factor on each side in order to obtain various dynamical quantities. Thus the equation $\ddot{x} = g$ becomes $m\ddot{x} = mg$. The quantity mg thus represents the force with which the earth is attracting the mass m—*i.e.*, the weight of the body. Hence, if w be the weight, we have the equation

$$w = mg,$$

which expresses the fundamental distinction between *weight* and *mass*. The mass m is fixed in amount; the weight w varies when g varies, and might be caused to vanish by taking the body to a region where g was zero.

It is frequently convenient to use, instead of m, the equivalent quantity $V\rho$, where V is the volume of the mass m and ρ is the mass per unit volume, which is called the *density* of the substance.

Again, the equation $\dot{x} = gt$ becomes $m\dot{x} = mgt$. But $m\dot{x}$ is the momentum produced in the body, which we thus see to be proportional to the time during which the body has been falling under the action of gravity.

Also, instead of $\dot{x}^2 = V^2 - 2gx$, we have $\frac{1}{2}m\dot{x}^2 = \frac{1}{2}mV^2 - mgx$, or $\frac{1}{2}m(V^2 - \dot{x}^2) = mgx$, which tells us that the loss of kinetic energy is proportional to the distance through which the body has risen.

In § 43 it was shown that the acceleration of a point moving in a circle of radius r with uniform speed v is v^2/r towards the centre. If the point have mass m, this corresponds to a central force mv^2/r, which, *e.g.*, in the case of a stone revolving in a sling, is supplied

by the tension of the cord. The necessity for this central force gave rise to the erroneous idea of a 'centrifugal force,' which it was supposed to balance. In accordance with the First Law, the body tends to move *along a tangent*, and not from the centre, and the apparent force is really a result of inertia.

65. *Further Discussion of the Second Law.*—The above examples involve the application of a single force, constant in magnitude and direction, to a material particle. But the Second Law enables us also to investigate the motion of a material particle under the action of any number of forces acting simultaneously, for it impilcitly asserts that each force acts independently of all the others, *i.e.*, the effect produced by any force is the same as it would be if that force alone acted upon the particle when at rest.

To completely specify a force we require to know its magnitude, the direction in which it acts, and the place at which it is applied. Hence a force is a vector quantity, and so the resultant of any number of forces acting simultaneously upon a material point is to be found by the ordinary law for the composition of vectors. Indeed this follows at once from the Second Law, since the forces are proportional to the accelerations which they produce.

Hence, when a particle is acted upon by any number of forces, we need only consider it as moving under the action of the single resultant force.

Since a mere particle has only three degrees of freedom—all translational—three conditions completely determine its motion. If X, Y, Z be the components of the resultant force in the directions of the axes of x, y, z, respectively, these conditions are, by the Second Law,

$$m\ddot{x}=\mathrm{X}\,;\ m\ddot{y}=\mathrm{Y}\,;\ m\ddot{z}=\mathrm{Z}.$$

In particular, the conditions of equilibrium are

$$\mathrm{X}=0\,;\ \mathrm{Y}=0\,;\ \mathrm{Z}=0.$$

66. *Special Examples.* — We shall now apply these results to some special cases of motion of a material particle.

(1.) A particle of mass m is projected with initial speed v_0. It experiences a resistance which is proportional to its velocity. Investigate the motion.

We may suppose that the motion is in the direction of the axis of x, so that we have only to consider the single equation

$$m\ddot{x}=-k\dot{x},$$

where k is constant. Multiplying each side by dt, this becomes

$$m\dot{x}=-kx.$$

The integral is

$$m\dot{x}=c-kx.$$

To determine the value of the constant c, we observe that v_o is the (given) value of $\dot{x}$ when $x=o$. Hence $c=mv_o$. Therefore, finally,

$$m(v_o-\dot{x})=kx.$$

This result shows that the particle will come to rest at a distance mv_o/k from the point of projection.

We may write the equation in the form

$$\frac{mdx}{mv_o-kx}=dt.$$

This gives (§ 38) $t=\mathrm{T}_o-m/k \log (mv_o-kx)$, and so, since $t=o$ when $x=o$, we have $\mathrm{T}_o=m/k \log mv_o$. But when $t=\mathrm{T}_o$ we get $x=mv_o/k$. That is $\mathrm{T}_o=m/k \log mv_o$ is the value of the time which elapses until the particle comes to rest.

(2.) A particle slides from rest down an inclined plane under the action of gravity. How long will it take to move over a given distance, and what will be its speed of motion when it reaches the given point?

Let m, α, R, and F represent respectively the mass of the particle, the inclination of the plane to the horizon, the normal pressure, and the force of friction. We may take the axis of x in the direction of motion, so that we get

$$m\ddot{x}=mg \sin \alpha-\mathrm{F}.$$

If μ' is the co-efficient of friction, this becomes

$$m\ddot{x}=mg \sin \alpha-\mu'\mathrm{R}.$$

If the axis of y be taken perpendicular to the plane, the other equation is

$$m\ddot{y}=0=\mathrm{R}-mg \cos \alpha.$$

Hence

$$\ddot{x}=g (\sin \alpha-\mu' \cos \alpha)=\mathrm{A} \text{ (say)}.$$

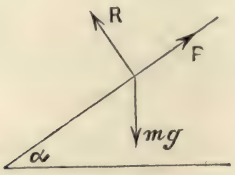


FIG. 47.

This quantity is independent of the mass of the particle. Multiplying by dt we get $d\dot{x}=\mathrm{A}dt$, from which

$$\dot{x}=v_o+\mathrm{A}t,$$

where v_0 is the initial speed, and therefore is zero by the given conditions. Hence, again,

$$x = x_0 + \tfrac{1}{2}At^2.$$

If we take the origin at the point from which the particle starts, we get $x_0 = 0$, so that the time taken to move over the distance x is

$$t = \sqrt{\frac{2x}{g\ (\sin\ a - \mu'\ \cos\ a)}}.$$

And the speed attained is

$$\dot{x} = \sqrt{2xg\ (\sin\ a - \mu'\ \cos\ a)}.$$

The kinetic energy gained by the particle is therefore

$$mgx\ (\sin\ a - \mu'\ \cos\ a).$$

(3.) A material particle is attached by an elastic cord to a point on an inclined plane down which it would slide under the action of gravity if not so attached. Find the limiting values of the tension in the cord between which motion will not occur.

T being the tension in the cord, and the other quantities having the same meaning as in (2), we get

$$m\ddot{x} = 0 = mg\ \sin\ a - \mathrm{T} \pm \mu mg\ \cos\ a.$$

From this equation the two values of T may be found. The sign + corresponds to the case in which the cord has its greatest extension so that friction acts down the plane. The sign − indicates that the particle is just on the point of sliding down, T having its smallest possible value.

(4.) A particle of mass m is swung round in a vertical circle by means of a cord of length l. What must be its angular velocity in order that the string may just be slack when the particle is at the highest point of its path?

The downward force is the weight of the particle, which must be balanced by the reaction to acceleration (the so-called 'centrifugal force'). Hence, at the highest point,

$$g = \omega^2 l,$$

where ω is the angular velocity.

At the lowest point we have

$$\mathrm{T} = mg + m\omega^2 l.$$

67. *Dynamical Similarity.* — We may write the expression for Newton's Second Law (§ 64) in the form

$$f = m\frac{l}{t^2},$$

where l is the distance which the mass m moves over from rest, in the time t, under the action of the force f. And we may further regard this equation as a dimensional equation (§ 27); in which case the sign of equality merely means that the dimensions of the quantity on the left-hand side of the equation are identical with those of the expression on the right-hand side of the equation. But, from a dimensional equation, we cannot make any deduction regarding the absolute magnitude of any of the quantities which are involved, for the equation simply asserts *proportionality* of magnitude between its various terms. Still, by a suitable definition of units, we can pass from the dimensional to the ordinary equation. Thus, in the above equation, we may define unit force as the force which, acting on unit mass for unit time, causes the unit of mass to move over unit distance from rest; or we might adopt the definition of § 64.

The idea of dimensions is of great importance in physics. It affords a useful check on the accuracy of algebraical work; for the dimensions of all the terms in a physical equation must be the same. But its use is not limited to this extent. For example, we may write the equation

$$f = m\frac{l}{t^2}$$

in the form

$$a\,f = \beta m \frac{\gamma l}{\frac{\beta}{a}\gamma t^2},$$

from which we see that *if, in two similar material systems, the forces, masses, and lengths, are in the ratios* $a/1$, $\beta/1$, *and* $\gamma/1$, *respectively, and if the systems begin to move in precisely similar manners, the motions will continue to be similar, provided that we compare them after the lapse of intervals of time which are in the ratio of* $\sqrt{\frac{\beta\gamma}{a}}$ *to unity in the two systems.*

This principle, which was proved otherwise by Newton, has been called the *Principle of Dynamical Similarity*. Later on, we shall get various examples of its use. (See §§ 73, 76, 124, 159.)

68. *The Third Law.*—Hitherto we have not discussed the motion of portions of matter between which there is mutual action of any kind. The first two laws of motion do not enable us to solve such problems. The requisite additional information is given by the Third Law of Motion: *The mutual actions between any two bodies are equal and oppositely directed.*

A *stress* is defined as a system of equilibrating forces, and so we may put the above law into the form: *The mutual action between any two bodies is of the nature of a stress.*

No one will question the truth of this law in the cases in which the various masses concerned are in equilibrium. Thus, when a book lies upon a table, we say that the table reacts upon the book

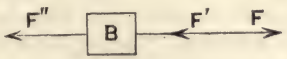


FIG. 48.

with a pressure which is equal to its weight. But it is by no means so evident that a body, when pulled along by means of a cord, pulls backwards with a force which is precisely equal to that by which it is dragged forwards. In order to see how this can be we must consider all the forces which are acting upon the moving body. Let B, Fig. 48, be the body and let F be the force with which it is pulled in the direction of F′F. Also let F″ be some other force acting upon B in the opposite direction.

The equilibrium of B is determined solely by the equality of the forces acting upon it, *i.e.*, of the forces F and F″, and is not at all affected by the force F′ with which B reacts upon the pulling body. Hence the equality of the forces F′ and F is a matter which is entirely independent of the equality of F″ and F, and can only be proved by experiment. It is needless to add that all Newton's laws express the results of experiment or of observation.

But (as Newton himself pointed out) we may regard 'action,' not merely as force but, as the product of force into the speed which it produces in the body upon which it acts. Now, the speed produced being the (time) rate at which the force moves the body, this product is (§ 62) the (time) rate at which work is done by the force. [In modern terminology this is called the *Activity*.] Hence a second interpretation of the third law is that *the (time) rate at which a set of forces do work upon a given system is equal and opposite to the rate at which the reacting forces do work.* Had Newton been aware that heat was a form of energy this would (see Thomson and Tait's *Elements of Natural Philosophy*) have been a complete statement of the modern principle of conservation of energy; but, in his day, it was supposed that work spent in overcoming friction is unavoidably and entirely lost.

Taken in conjunction with the Second Law, this law enables us to investigate the motion of bodies which impinge upon each other.

To avoid unnecessary complications we may assume that two smooth spheres, of masses m_1 and m_2 respectively, are moving in the direction of the line joining their centres with speeds v_1 and v_2 respectively, and that after impact their velocities are v'_1 and v'_2. In most practical cases the time of impact is a very small fraction of a second, and the force is very large, so that it is impossible, without special appliances, to determine the values of these quantities. But the value of their product, called the *Impulse*, can generally be found without much difficulty. The third law tells us that the impulse is the same for each body, and hence

$$m_1(v'_1 - v_1) = m_2(v_2 - v'_2).$$

In addition to this we have the condition, determined experimentally by Newton,

$$v'_1 - v'_2 = e(v_2 - v_1),$$

where e is a constant, less than unity, which is called the *Coefficient of Restitution*. This condition asserts that the relative speed of separation of the two bodies is less than, but is proportional to, their relative speed of approach. It ceases to be true if the distortion produced by the impact is too great.

If the bodies, after impact, move together with a common speed V, the first of these equations becomes

$$(m_1 + m_2)\mathrm{V} = m_1v_1 + m_2v_2.$$

This principle is employed in the *Ballistic Pendulum*, which is used to determine the speed of a cannon ball or of a rifle bullet. In this case the mass of the pendulum, m_2, is very large in comparison with the mass m_1 of the bullet, and v_2 is zero. The large relative value of m_2 ensures that the two masses are moving together with the common speed V before the pendulum has been sensibly deflected from the vertical. The value of V is found by observing the distance through which the pendulum swings. [From this, the height h through which the centre of inertia (§ 69) is raised is obtained, and then (§ 42) we get $\mathrm{V} = \sqrt{2gh}$.]

69. *Centre of Inertia.*—In a material system composed of masses m_1, m_2, etc., we can always find a point such that the product of its distance from any plane into the sum of the separate masses is equal to the sum of the products of each separate mass into its own distance from that plane.

Let $\Sigma(m)$ denote the sum of the masses, and let $\Sigma(md)$ denote the sum of the products of each mass into its distance d from *one* given plane. We then can obviously find the distance D from this plane such that

$$\Sigma(m)\mathrm{D} = \Sigma(md), \ldots\ldots\ldots (a),$$

for this is a single equation in one unknown quantity D. Let this be done for other two planes, neither pair of the three being parallel, and we get a fixed point, which satisfies the condition for these three planes.

Let us suppose, for convenience, that the three planes are at right angles to each other, and that the lines of intersection are taken as the axes of x, y, and z respectively. We have then the three equations similar to the above,

$$\Sigma(m)\mathrm{X}=\Sigma(mx),$$
$$\Sigma(m)\mathrm{Y}=\Sigma(my),$$
$$\Sigma(m)\mathrm{Z}=\Sigma(mz).$$

Now, multiplying these equations respectively by any quantities λ, μ, ν, and adding, we get

$$\Sigma(m)(\lambda\mathrm{X}+\mu\mathrm{Y}+\nu\mathrm{Z})=\Sigma(m(\lambda x+\mu y+\nu z)).$$

But λ, μ, and ν may be the direction-cosines of the normal to any plane passing through the intersection of the three given planes, in which case the quantity $\lambda x+\mu y+\nu z$ is the perpendicular upon this plane from any point whose co-ordinates are x, y, z. Hence equation (a) is true for any plane which passes through the intersection of the three given planes.

But equation (a) is still satisfied if we increase D and d by any constant quantity h, for this simply adds on $\Sigma(m)h$ to each side; that is to say, (a) holds for any plane parallel to a given one for which it is true. Hence, it holds for all planes.

The point so found is called the *Centre of Inertia* of the given set of material particles.

By taking the time-rate of variation of the quantities in the above equation we obtain

$$\Sigma(m)\dot{\mathrm{D}}=\Sigma(m\dot{d})$$

and

$$\Sigma(m)\ddot{\mathrm{D}}=\Sigma(m\ddot{d}).$$

The former tells us that the momentum of the system in any given direction is equal to the momentum, in that direction, of a single mass, equal to the sum of the separate masses, moving so as always to be situated at the centre of inertia. The latter asserts that the change of motion of the centre of inertia of any set of disconnected particles produced by the action of separate forces on the separate masses is the same as if these forces had been applied to a mass, equal to the total mass, condensed at the centre of inertia.

In consequence of the equality of action and reaction between

material particles, we see that the motion of the centre of inertia of any connected set of particles is not affected by their mutual action; and that, in the case of a rigid body, we may suppose the whole mass to be condensed at the centre of inertia, and to be acted upon by the resultant force. In other words, the equations of translational motion and equilibrium of a rigid body may be made identical with those already given in § 65 for a material particle.

70. *Moment of a Force and of Inertia.*—The *Moment of a Force* as regards rotation about an axis perpendicular to its direction is the product of the force into the shortest distance between its line of action and the axis.

A pair of parallel, equal, and oppositely directed, forces is called a *couple*. The moment of a couple about any axis perpendicular to the plane in which the forces act is equal to the product of either force into the perpendicular distance between the lines of action of the two. Let r be this distance, and let F be the common value of the forces, while P is the intersection of any perpendicular axis with

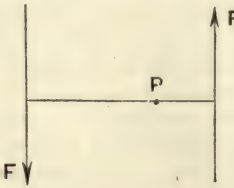


FIG. 49.

the plane in which the forces act. If x is the perpendicular distance from P to the line of action of one force, $r-x$ is the perpendicular distance from P to the line of action of the other. Hence, the sum of the moments of the forces, which is the moment of the couple, is

$$\mathrm{F}x+\mathrm{F}\ (r-x)=\mathrm{F}r.$$

Now we may write $\mathrm{F}r=mar$, where m is the mass acted upon, and a is the linear acceleration. But $a=\omega r$, where ω is the angular acceleration. Hence

$$\mathrm{F}r=mr^2\omega.$$

Three independent equations of this type completely specify the rotational motion of the given mass.

The quantity mr^2 is called the *Moment of Inertia* of the mass m about the given axis. Multiplying each side of the equation by ωdt and forming the integrals, we get

$$\mathrm{F}r\theta=\tfrac{1}{2}mr^2\omega^2,$$

if $d\theta/dt=\omega$, so that θ is the whole angle through which the mass has turned. But since $r\omega=v$, the linear speed, the quantity on the right-hand side of the equation is the kinetic energy of rotation. Hence, we see that the kinetic energy of rotation acquired under the action of a given couple is proportional to the angle through which the mass has turned.

If we are not dealing with a single particle, we must write the sum $\Sigma(mr^2)$ instead of mr^2 in the above equation. But we may still put the equation in the same form as before by writing

$$\Sigma(m)k^2=\Sigma(mr^2),$$

which is clearly allowable, since k (called the *radius of gyration*) is the only unknown quantity.

As an example, we shall investigate the motion of a cylinder rolling (not sliding) down a plane inclined at an angle α to the horizon. Let r be the radius of the cylinder, and let k be its radius of gyration. The distance through which the cylinder descends when it turns through an angle θ, is $s=r\theta \sin \alpha$. If m is the mass

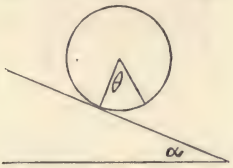


FIG. 50.

of the cylinder, the work done by gravity is $mgs=mgr\theta \sin \alpha$. This must be equal to the gain of kinetic energy. The energy in the rotational form is $\frac{1}{2}mk^2\dot{\theta}^2$, and that in the translational form is $\frac{1}{2}m(r\dot{\theta})^2$. Hence, $v(=r\dot{\theta})$ being the speed of motion down the plane,

$$2gs=(r^2+k^2)\dot{\theta}^2=\frac{r^2+k^2}{r^2}v^2.$$

Had there been no rotation, we should have had $v^2=2gs$; but the speed of linear motion has been decreased because the potential energy became transformed in part into energy of rotation.

71. *Further Discussion of Moment of Inertia.*—The moment of inertia of a body about any axis is equal to its moment of inertia about a parallel axis through the centre of inertia, together with the moment of inertia, about the original axis, of a mass, equal to the whole mass, condensed at the centre of inertia.

The moment of inertia is

$$\Sigma(mr^2)=\Sigma(m\overline{x^2+y^2}).$$

Transfer the origin to P (Fig. 51), the centre of inertia, the co-ordinates of which are a, β. Let ξ, η be the co-ordinates of any point referred to parallel axes through P. Then

$$\Sigma(m\overline{x^2+y^2}) = \Sigma(m\overline{(a+\xi)^2+(\beta+\eta)^2}) = \Sigma(m)\overline{a^2+\beta^2} + \Sigma(m\overline{\xi^2+\eta^2}),$$

since $\Sigma(m\xi)$, $\Sigma(m\eta)$ vanish by the properties of the centre of inertia.

To illustrate the importance of this result we shall proceed to find the moment of inertia of a cylindrical rod, of length $2l$ and radius a, about an axis drawn through its centre perpendicular to its length. Consider a circular disc of the rod of infinitesimally small thickness dh. The moment of inertia of this disc about the axis of the rod is

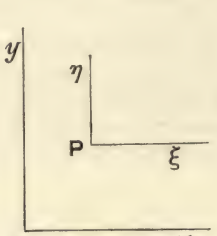


Fig. 51.

$\Sigma(mr^2)$, where r is the distance of the elementary mass m from the axis, and the summation extends from $r=o$ to $r=a$. If ρ is the density of the rod, we may write $2\pi r dr\rho dh$ instead of m; for $2\pi r dr$ is the area of a small circular ring of the disc, so that $2\pi\rho r dr dh$ is the mass of a small annular portion of the disc. The moment of inertia of this part is therefore $2\pi\rho dh r^3 dr$, and the integral of this from $r=o$ to $r=a$ is the moment of the whole disc. It is therefore $\frac{1}{2}\pi\rho dh a^4$.

Now, since the moment of the whole disc about a central perpendicular axis is

$$\Sigma(mr^2) = \Sigma(m(x^2+y^2)) = \Sigma(mx^2) + \Sigma(my^2),$$

where x and y are the co-ordinates of m referred to any two mutually rectangular central axes in the plane of the disc, and since $\Sigma(mx^2)$ and $\Sigma(my^2)$ are respectively the moments of inertia of the disc about the axes of x and y, we see that the moment of the disc (or of any plane figure) about an axis in its own plane, drawn through its centre of inertia, is one-half of its moment about a perpendicular axis through its centre of inertia.

The moment of the disc about a central axis in its plane is therefore $\frac{1}{4}\pi\rho dh a^4$. And, if h be the distance of the centre of the disc

from the centre of the rod, the moment of inertia of the whole mass ($\pi\rho a^2 dh$) of the disc, supposed to be condensed at its centre, about a line passing through the centre of the rod and perpendicular to its length, is $\pi\rho a^2 dh h^2$. Hence the moment of inertia of the disc about this line is $\pi\rho a^2\left(\frac{a^2}{4}+h^2\right)dh$. And, if we sum the moments of all such discs from $h=o$ to $h=l$, we get half the moment of inertia of the whole rod. Taking twice the integral of this quantity between these limits, we find that the required moment is $2\pi\rho l a^2\left(\frac{a^2}{4}+\frac{l^2}{3}\right)$; that is, $M\left(\frac{l^2}{3}+\frac{a^2}{4}\right)$, where M is the whole mass of the rod.

72. *Rotational Equilibrium.*—There is no rotation about a given axis when $\Sigma(Fr)$, taken with reference to that axis, is zero. Hence, the condition for rotational equilibrium is that the sum of the moments of all the forces about three non-parallel axes shall vanish. The two following examples will serve to illustrate this point.

(1) A uniform ladder (Fig. 52), of length $2l$, rests, in a vertical plane, upon the ground, and a vertical wall. Find the limiting posi-

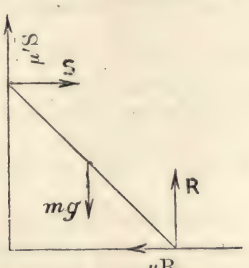


FIG. 52.

tion of equilibrium, the co-efficients of friction between the ladder and the ground, and between the ladder and the wall, being μ and μ' respectively.

Let a be the inclination of the ladder to the ground. Under the given conditions, there is no possibility of motion except in the given vertical plane. Hence, there are only three degrees of freedom, viz., two degrees as regards translation, and one as regards rotation. The ladder will be in equilibrium, so far as translation is concerned, if the sums of the forces acting upon it in *any* two mutually perpendicular directions are zero. But it is convenient to

choose those two directions which will lead to the simplest equations. We might choose the directions along and perpendicular to the length of the ladder, but each of the consequent equations would involve all the five forces which are acting. If we choose the horizontal and vertical directions, the equations involve respectively two and three forces only. Therefore, choosing the latter directions, we get

$$S = \mu R,$$
$$mg = R + \mu' S,$$

where S and R are the normal pressures on the wall and the ground respectively.

The third relation between the quantities is obtained by equating to zero the sum of the moments of the various forces about any axis perpendicular to the plane of motion. The simplest equation is obtained by choosing the axis passing through a point (either end of the ladder), which lies on the lines of action of the greatest possible number of forces; the reason being that the moments of these forces are then zero. Taking the lower end we get

$$mgl \cos \alpha = (S \sin \alpha + \mu' S \cos \alpha) 2l,$$

that is

$$mg \cos \alpha = 2S(\sin \alpha + \mu' \cos \alpha),$$

which is independent of the length of the ladder.

If we suppose the weight of the ladder, and the values of μ and μ' to be given, we may eliminate, by means of these three equations, the quantities S and R, and so obtain an equation giving α in terms of known quantities.

(2) A pendulum, of length l and mass m, rotates about a vertical axis with constant angular velocity ω. Express ω in terms of g, the value of gravity, and of h, the height of the cone which the pendulum describes.

Let θ be the angle which the pendulum makes with the vertical. The 'centrifugal force' acting perpendicular to the axis is $m\omega^2 l \sin \theta$. The portion of this which is perpendicular to the string of the pendulum, and which acts so as to prevent decrease of θ, is $m\omega^2 l \sin \theta \cos \theta$. The part of the weight which acts in the same line, but inwards so as to decrease θ, is $mg \sin \theta$. Hence the condition for equilibrium is

$$\omega^2 l \cos \theta = \omega^2 h = g.$$

This gives the required expression for ω, and shows that the angle θ is constant.

73. *Propagation of Motion through a Non-Rigid Solid.*—As a

single example of the motion of a non-rigid solid we shall now discuss the problem of the passage of a wave along a stretched cord.

Let us suppose the cord to be enclosed in a smooth hollow tube, and to be drawn through it in the direction of the arrow with speed v.

FIG. 53.

The tension T of the cord will be uniform throughout since the tube is smooth. The pressure which the cord, if not in motion, would exert upon a part of the tube where the radius of curvature is r would be T/r. Let the cord be in contact throughout a circular arc PQ which subtends an angle θ at the centre O. Then, if OR bisects θ, the resolved part of the tensions along RO is $2T \sin \theta/2$. If θ is small this becomes $T\theta = T \cdot PQ/R$. But the total pressure is pPQ, where p is the pressure per unit length of the circle. Hence

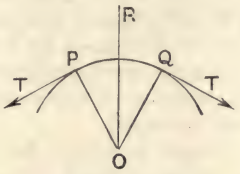


FIG. 54.

$p = T/R$; that is to say p is proportional conjointly to the tension and the curvature.

If m is the mass per unit length of the cord, mv^2/r is the 'centrifugal force' when the cord is in motion with speed v. When this is equal to T/r, *i.e.*, when $T = mv^2$, there is no pressure on the surface. The value of v which satisfies this equation is totally independent of r, the radius of curvature. Hence, when the proper speed is reached the pressure is simultaneously taken off all parts of the smooth tube through which the cord runs; and the tube, having served the purpose for which it was used, might now be dispensed with. All parts of the cord would successively take the shape which the tube originally impressed upon the portion within it. And also, since all motion is relative, if the cord were held fixed with the given constant tension, the wave-form would run

backwards along it with speed v. Hence the speed with which any disturbance will run along a cord stretched with tension T is

$$v = \sqrt{\frac{T}{m}},$$

when m is the mass per unit length.

Simple as the above proof (due to Thomson and Tait) is, the following, deduced from the principle of dynamical similarity, is at least as simple.

The radius of curvature at similar parts of similar waves is proportional to the length l of the wave. The pressure per unit of length is therefore proportional to T/l, so that the pressure per similar length is proportional to T. Also the mass per similar length is ml, and hence we get the dimensional equation

$$T = \frac{ml^2}{t^2},$$

where t is the periodic time in which the wave length l is described. But l/t is the speed of propagation, which is therefore equal to $\sqrt{T}/\sqrt{m}$, if we adopt the definition of force given in § 64.

74. *Motion of a Perfect Fluid.*—A fluid may be set in motion by the action either of forces which act throughout its volume (for example, gravitational forces) or of forces which are applied to its surface (such as external pressure).

A *perfect* fluid is defined as a fluid in which the pressure is always perpendicular to the surfaces of contact. It may otherwise be defined as a fluid which is entirely devoid of internal friction. Such a fluid does not exist in nature, but we may deduce various results regarding the motion of perfect fluids which will be very nearly true for actual fluids which are moving with sufficient slowness. [When *any* fluid, whether perfect or not, is at rest, the pressure is always perpendicular to the surfaces of contact.]

Consider a little cube with edges dx, dy, dz, parallel to the axes in a fluid of density ρ. The mass of this little cube is $\rho dxdydz$, and its acceleration of momentum parallel to the x-axis is $\rho\ddot{x}dxdydz$; and this is equal to the sum of the forces acting upon the little mass in that direction. Let X be the force per unit of mass which is acting upon it, and let p be the pressure per unit of area. The total pressure on the face of the cube next the origin is $pdydz$, and this acts outwards along the x-axis. When x changes by the amount dx, p will alter to $p + dp$, so that the pressure acting inwards along

the axis is $(p+dp)dydz$. Hence the total *outward* pressure is $-dpdydz$, or as we may write it, $-\frac{dp}{dx}dxdydz$. Hence we have as the equation of motion parallel to the x-axis

$$\rho\ddot{x}=\rho X-\frac{dp}{dx}.$$

Similarly

$$\rho\ddot{y}=\rho Y-\frac{dp}{dy},$$

and

$$\rho\ddot{z}=\rho Z-\frac{dp}{dz}.$$

[It must be carefully noticed that $\ddot{x}$, $\ddot{y}$, and $\ddot{z}$, represent the *total* component accelerations, which may vary independently with the time and with the position of the small mass: in short, we are supposed to follow the given portion of the fluid in its motion.]

As an example we shall investigate the motion, under gravity, of a fluid which escapes through a small orifice in the side of a vessel, the depth of the opening below the free surface of the liquid being z. Take the origin at the free surface, the axis of z being drawn downwards, and the axes of x and y being horizontal. The equations of motion are

$$\rho\ddot{x}=-\frac{dp}{dx};\ \rho\ddot{y}=-\frac{dp}{dy};\ \rho\ddot{z}=\rho g-\frac{dp}{dz},$$

g being the acceleration due to gravity. Multiplying these equations by $\dot{x}dt$, $\dot{y}dt$, $\dot{z}dt$, respectively, and adding we get,

$$\rho(\dot{x}d\dot{x}+\dot{y}d\dot{y}+\dot{z}d\dot{z})=\rho gdz-\left(\frac{dp}{dx}dx+\frac{dp}{dy}dy+\frac{dp}{dz}dz\right)$$
$$=\rho gdz-dp.$$

The integral of this is

$$p=p_o+\rho gz-\rho\tfrac{1}{2}v^2, \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

where $v=\sqrt{\dot{x}^2+\dot{y}^2+\dot{z}^2}$ is the speed of motion of the fluid and p_o is the pressure on the free surface of the liquid since it is the value of p when $z=o$ and $v=o$, which is practically the case when the area of the opening by which the fluid escapes is very small in comparison with the free surface of the liquid.

The pressure just outside the opening is p_o, and hence we have

$v^2 = 2gz$. That is, the speed is that which would be acquired in a free fall from rest under gravity through a distance equal to the depth of the opening below the surface of the liquid.

This result may readily be deduced by considerations regarding the energy of the liquid. The kinetic energy of a quantity m of the the escaping liquid is $\frac{1}{2}mv^2$. But this energy, which the escaping liquid carries away with it, is at once restored if we simply pour the liquid back again into the vessel. And the work done in raising the liquid through the height z is mgz. Hence, by the principle of conservation of energy, we have, as before, $v^2 = 2gz$.

Equation (1) shows that, in a moving fluid, the pressure is least where the speed is greatest. Hence there is less pressure in the interior of a moving jet of fluid than there is at the outside. Thus objects immersed in the fluid will be pressed inwards to the centre of the jet. This explains the support of a light body in a vertical jet of water or of air.

75. *Equilibrium of a Fluid.*—From the equations of motion of a fluid we at once get, as a special case, the conditions of equilibrium. These are

$$\rho X = \frac{dp}{dx};\ \rho Y = \frac{dp}{dy};\ \rho Z = \frac{dp}{dz}.$$

If no external forces act upon the liquid, we have the equations

$$\frac{dp}{dx} = 0;\ \frac{dp}{dy} = 0;\ \frac{dp}{dz} = 0,$$

which simply assert that the pressure has a constant value at all points of the liquid.

If we assume that the origin is at the surface of the liquid, that gravity acts, and that the axis of z is drawn downwards, the equations are

$$\frac{dp}{dx} = 0;\ \frac{dp}{dy} = 0;\ \frac{dp}{dz} = \rho g.$$

The first two assert that the pressure is constant throughout a horizontal plane, and the last shows that it increases uniformly with the depth. The integral is

$$p = p_0 + \rho g z.$$

This might have been obtained from (1) of last section by making $v = 0$.

76. *Propagation of Surface-Waves in Liquids.*—We shall assume,

for the sake of simplicity, that the waves are all similar, and that their ridges are parallel equi-distant straight lines. The principle of dynamical similarity then enables us to deduce easily the law of propagation.

When the waves are propagated by gravity, the forces are proportional to the density of the liquid, to the value of gravity, and to the square of the wave-length; and the masses are proportional to the density, and to the square of the wave-length. Hence the dimensional equation

$$f = ml/t^2,$$

(§ 67), becomes

$$\rho l^2 g = \frac{\rho l^2 \cdot l}{t^2}$$

or

$$g = l/t^2,$$

which gives

$$v^2 = gl.$$

In these equations ρ represents the density of the liquid, and the other symbols have the usual significations. We see, therefore, that the speed of propagation is proportional conjointly to the square roots of the wave-length, and of the acceleration due to gravity. Such waves are called *oscillatory* or *free* waves.

In the above case, it is assumed that the depth of the liquid is very large in comparison with the length of the waves. When the depth is very small in comparison with the wave-length, the above equations still apply, provided that l represents the depth of the liquid. For, when similar waves are propagated in liquids of different depths (the similarity having reference to the depth), we see that similar masses are proportional to the squares of the depths, while the ranges of vertical motion are proportional directly to the depth. Hence the speed of propagation of such waves, which are called *long* or *solitary* waves, is proportional conjointly to the square roots of the depth and of the acceleration due to gravity.

In the propagation of *ripples*, surface-tension (Chap. X.) is much more effective than gravitation is. If T represents the surface-tension, while l represents the wave-length, the pressure per unit area of the surface is proportional to T/l. Hence the pressure per similar area is proportional to T, for we are not concerned with lengths measured parallel to the ridges of the waves. The similar masses are proportional to ρ and to l^2, and so we get

$$v^2 = \frac{T}{\rho l}.$$

The speed of propagation of a ripple is therefore proportional

directly to the square root of the surface-tension, and inversely to the square root of the product of the density into the wave-length. Thus we see that ripples run faster the smaller they are, while oscillatory waves run faster the larger they are. Hence there is a certain size of wave (about two-thirds of an inch in length in the case of water) which runs slowest. Smaller waves run more quickly because the effect of surface-tension preponderates; larger waves run more quickly because of the increased effect of gravity.

CHAPTER VII.

PROPERTIES OF MATTER.

77. *Definitions of Matter.*—We are now in a position to give one or two provisional definitions of matter—provisional, because we cannot yet say, possibly may never be able to say, what matter really is. It may be defined in terms of any of its distinctive characteristics. We may say that *Matter is that which possesses Inertia.* Or again, since we have no knowledge of energy except in association with matter, we may assert that *Matter is the Vehicle of Energy.* Another statement (which, from the results of Chap. I. and § 62, we see to be an objectionable form of the latter) is that 'matter is that which exerts, or can be acted upon, by force.' Further knowledge would probably make it evident that these three definitions are merely differently worded statements of the same fact.

78. *States of Matter.*—Matter is usually spoken of as existing in three different states—the solid, the liquid, and the gaseous.

A portion of matter in the solid state possesses a definite form of its own, and considerable force has to be applied in order to produce an appreciable change in the form. When in the liquid state, matter, on the other hand, possesses no definite form of its own, and can be made by application of the slightest force to change whatever form it happens to have. A similar statement holds in the case of a gas or vapour. But a gaseous body differs from a liquid in that its volume is limited only by the volume of the closed vessel in which it is contained; while the volume of a given quantity of liquid is perfectly definite, under given physical conditions, however large the containing vessel may be.

Still—although these rules are of general applicability—it must not be supposed that there is any hard and fast distinction between a solid and a liquid, or between a liquid and a vapour: possibly (but there is no experimental proof of the truth of this statement) there may be,

under certain physical conditions, no sharp line of demarcation between the solid and the vaporous states of matter.

As the temperature of sealing-wax is gradually raised, the substance slowly passes from the solid into the liquid condition, and, for some time, we cannot strictly call it either a liquid or a solid. The transition from ice to water seems to occur suddenly, but analogy would indicate that the process is really a continuous one. We shall also find (Chap. XXIII.) that the passage from the vaporous to the liquid condition may be made without break of continuity.

[A very notable and extremely important example of the non-rigidity of the above distinctions occurs in the case of shoemakers' wax. This substance so far resembles a brittle solid that it will break into splinters under the blow of a hammer; and yet, under the action of slight long-continued forces, it can be moulded into any shape we please.]

The extreme form of the gaseous condition, which is known as the 'ultra-gaseous' or 'radiant' state of matter, will be discussed in Chap. XIII.

79. *General Properties.*—Certain properties are common to all portions of matter in whatever state or physical condition they may be, and are, therefore, called *general* properties.

Chief among them is the already-mentioned property of *inertia*, which requires no further discussion at present.

Again, all matter occupies space; and we consequently say that it has the property of *extension*. This subject has been considered in Chap. III. But the occupancy of space further involves the idea of form, and so we recognise *form* as a property of matter. There is little more to be said on this point except in the case of the form of crystallised bodies, to the consideration of which a considerable part of Chap. XII. will be devoted.

So far as we know any one portion of matter occupies a given portion of space to the utter exclusion of all other matter. This certainly holds in the case of any visible portion. Hence we look upon *impenetrability* as a general property of matter. But the quality of impenetrability does not interfere with inter-penetration of matter, the possibility of which depends upon the existence of pores in any finite portion of matter. The corresponding property is called *porosity*. All matter is more or less sponge-like in structure, the space in the so-called 'internal' pores being in reality external to the material of the body. In the case of substances such as cork, wood, coarse sandstone, etc., the porosity is very evident. The porosity of metals is shown by the fact that gases can pass through them. Thus palladium has a remarkable power of absorbing or

'occluding' hydrogen; carbonic oxide passes readily through red-hot iron; and gases formed by the decomposition of electrolytes pass through the metallic electrodes. Bichromate of potassium passes into the pores of glazed earthenware, and, crystallizing inside, gradually breaks up the substance. The porosity of liquids is evidenced by their absorption of gases.

Vitreous bodies alone have not yet been directly shown to be porous; but we may fairly conclude that we have not yet found the proper method of testing the point, and that, the proper method being found, they too will prove to be no exception to the rule.

A very noteworthy example of interpenetration occurs in the alloying of certain metals, such as tin and copper. The bulk of this alloy is considerably less than the sum of the bulks of its constituents. This phenomenon does not occur when gold and silver are alloyed, and for this reason only was Archimedes' famous test of the impurity of Hiero's crown conclusive. A certain weight of pure gold had been given to a smith for the purpose of making the crown; but it was suspected that he had abstracted some of the gold, replacing it by an equal weight of silver. Archimedes knew that, weight for weight, silver is bulkier than gold; and hence he concluded that the crown would be bulkier than the given amount of pure gold if silver had been used as an alloy. The problem which he required to solve was therefore the determination of the bulk of a solid which was so irregular in shape that no method of estimation by direct measurement was applicable. He determined this by measuring the volume of the water which the crown displaced. Had contraction taken place, the bulk and weight of the alloy might have been the same as those of the pure gold.

Another property of all matter is *divisibility*. The question of the infinite divisibility of matter will be further alluded to in the chapter on the constitution of matter. In the meantime we are only concerned with examples of extreme division. Many such occur readily. Films of gold and of other metals may be made so thin as to be transparent. A film of gold precipitated by chemical means and burnished so that it forms a continuous sheet may be of no greater thickness than one ten-millionth part of an inch. A quartz-fibre may be made so fine as to be utterly invisible. The vapour from a particle of sodium will tinge a flame continuously of a deep orange colour for hours at a time. A single drop of a strongly coloured liquid will continuously tinge a very large quantity of water, and its presence may be made evident by chemical means long after the eye ceases to detect it. Further examples of the

extreme smallness of portions of matter will be given in the chapter on the constitution of matter.

All matter is capable of having its volume diminished under pressure to a greater or less extent, and so we speak of its *compressibility*. This subject will receive detailed treatment in Chaps. IX., X., and XI.

Again, all matter is deformable. But it is often more convenient to speak of its *rigidity* than of its deformability, that is, the property in virtue of which it resists deformation. This question, too, will be discussed subsequently.

We shall also afterwards consider more specially *elasticity*, which, in one or other of its two forms, exists in all kinds of matter; and *viscosity*, that is, the property in virtue of which there is resistance to relative motion of the particles of a body. *Expansibility* will be dealt with in Chap. XXII.

Weight—which, though a universal property of matter, may be looked upon as a purely accidental property, seeing that it requires the existence of two separate masses in order that it may appear—will be fully considered in the chapter on 'Gravitation.' (See also Chap. VI., § 64.)

80. *Special Properties.*—Many other properties might be enumerated which are conspicuously present in some substances, and are as conspicuously absent from others, such as *plasticity*, *ductility*, *brittleness*, *tenacity*, etc. Again, many properties refer to matter in connection with special forms of energy; for example, *dispersive power*, *thermal* and *electric conductivity*, *magnetic permeability*, *translucency*, *opacity*, etc. Indeed, all the properties of matter might be naturally investigated in a treatise on energy; for we have no notion of what might be the properties of matter devoid of energy. Most probably it would not be matter at all in the sense in which we use the word.

81. *Specific Properties.*—Many of the properties of a body depend upon the size of that body. Thus a portion of a given substance is more massive than another portion of the same substance in proportion as its bulk is greater than the bulk of that other.

It is frequently very essential to define a property of a body in such a way as to make it independent of the size of any particular specimen. For example, the *density* (specific mass) of a substance is the mass per unit volume of that substance; the *specific gravity* is the weight of unit volume, expressed in terms of that of water as the standard; the *specific weight* might be defined as the weight per unit volume, expressed in absolute units. We also define *rigidity*, *viscosity*, etc. (§§ 108, 128) as specific properties.

82. *Molecules and Atoms.*—A visible portion of any chemically compound substance, *e.g.*, water, may be divided into smaller parts each of which has the same chemical peculiarities as the whole had. These parts may even be so small as to be invisible to the most powerful microscope. But at last a stage is reached where further division cannot occur without the production of substances which are chemically different from the original one. These smallest parts, which are chemically similar to the whole, are called *molecules.*

Every molecule can be divided further into what are termed its constituent *atoms.* These atoms are dissimilar when the molecule is complex like that of water: they are all precisely similar in a simple molecule, such as that of hydrogen. An atom is an indivisible part—that is, indivisible by any means at present at our disposal. In all probability, it is really compound.

A substance whose molecules are composed entirely of similar atoms is called an 'elementary' substance, and the kind of matter composing it is called a 'chemical element.' We know of the existence in nature of only a comparatively small number of elements.

One fact of the utmost importance is this—that *wherever in space it may be situated, or howsoever it may be circumstanced physically, an elementary molecule or atom has absolutely unalterable properties.* A molecule of hydrogen (for example) in the most distant nebula is precisely similar to a molecule of that substance on the earth's surface. We shall return to this point later on.

83. *Molecular Forces.*—A considerable, frequently a very great, amount of work is necessary in order to separate the molecules of a body. When separated, we are accustomed to say that they possess potential energy of molecular separation—the increase of potential energy being equivalent to the work spent in producing the separation. If e, the energy, increases by the amount de when the distance s between the molecules is increased by the quantity ds, the work spent is represented also by the quantity

$$de = \frac{de}{ds} ds.$$

The quantity de/ds—the space-rate of variation of the energy—is called the *molecular force* against which the work is done (see § 29). That is to say, we figure the molecules to ourselves as held together by certain forces which maintain them in their relative positions.

The work done in separating two molecules beyond the range of

their mutual forces is usually great, and practically the whole of it is performed in an excessively short distance. Hence de/ds is very large; and so it is said that the molecular forces are extremely powerful, but that they are insensible at sensible distances. In confirmation of this we may firmly press together two leaden bullets at a part where their surfaces have been freshly cleaned. They will cling together so that the lower one may be lifted by means of the upper, the magnitude of the so-called molecular forces at the surface of contact being sufficient to overcome the gravitational attraction of the whole earth. If there is a film of oxide on the metal the experiment will not succeed, the thickness of the film being too great to allow of appreciable molecular attraction. (For further statements on this subject, see Chap. XII.)

Many of the properties of matter may be said to depend on the nature of the molecular forces. *Tenacity* depends upon the extent to which these forces can overcome external forces which tend to draw the molecules apart. *Malleability* is a property essentially analogous to the preceding. It is the property in virtue of which a substance may be extended in two directions, while it is contracted in a direction perpendicular to these by the application of great pressure, its volume remaining practically unaltered. In testing tenacity the extension occurs in one direction only, and there is contraction in all directions perpendicular to that one. The *brittleness* of a body is due to the comparative ease with which the molecules can be separated beyond the range at which their mutual forces are appreciable. *Rigidity* depends upon the ability of the molecular forces to resist alteration of the relative positions of the molecules of a body. We therefore recognise two kinds of rigidity —rigidity as regards bulk, and rigidity as regards form. *Viscosity* depends upon the ability of the forces to resist shearing motion.

84. In the immediately succeeding chapters, a more detailed examination of some of the most important of the properties alluded to above will be given. Others will be treated as occasion may arise subsequently.

CHAPTER VIII.

GRAVITATION.

85. All bodies in the neighbourhood of the earth's surface possess potential energy, which, when circumstances permit, is invariably changed into kinetic energy of motion *towards* the earth. Hence we say that each body is 'attracted' to the earth with a force which is termed its 'weight.'

Bodies, made of the same material, may be roughly judged to be heavy in proportion to their bulk, *i.e.*, in proportion to the quantity of matter contained in them. But we need not rest content with a roughly approximate rule, for the law that weight is proportional to mass is capable of as rigid proof as can be given by the most accurate physical methods.

In the first place every body—except in so far as the resistance of the air is concerned—takes the same time to fall through a given distance. In other words, the acceleration of motion is the same

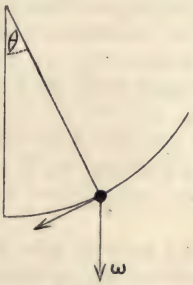


Fig. 55.

in all cases. Hence, by Newton's Second Law, the force (weight) is proportional to the mass.

The fact that the time of oscillation of a simple pendulum is independent of the mass of the bob furnishes a more readily obtainable

proof of the proportionality of weight and mass. Let θ be the angle through which the pendulum is deflected from the vertical. If w is the weight of the body the force acting in the direction of motion is $w \sin\theta$; which, for small angles, is practically $w\theta$. This must be equal to the acceleration of momentum, which is $ml\ddot{\theta}$, if l is the length of the pendulum. Hence $w\theta = ml\ddot{\theta}$; and therefore (§ 51)

$$\theta = a\cos\left(\sqrt{\frac{w}{lm}}t + \alpha\right),$$

a and α being constants. This shows that the motion is simple harmonic. And the time of a complete oscillation is

$$2\pi\sqrt{\frac{lm}{w}},$$

since the value of θ is unaltered if we increase t by this amount. Now experiment shows that this quantity is constant when l is constant. Hence w is proportional to m.

Again, a body has the same weight whether its surface is large or small, and whether it is in a single lump or broken up into parts. This proves that the outer parts do not screen the interior parts from the action of gravitation. Indeed, perpetual motion could ensue if this were not so; for, if matter were placed between one half of a vertically mounted wheel and the earth, the other half of the wheel would be permanently heavier.

86. The truth of Kepler's Laws regarding the motion of a planet would prove that the force of attraction between the planet and the (fixed) sun is in the direction of the line joining their centres, and is inversely proportional to the square of the distance between them. These laws are:—

I. Each planet moves in an ellipse of which the sun occupies one focus. (The path of a comet may be any conic section.)

II. In that ellipse the radius-vector traces out equal areas in equal times.

III. The square of the periodic time is proportional to the cube of the mean distance between the planet and the sun.

If equal areas are traced out in equal times, the quantity

$$\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta})$$

is zero (§ 50), which means that the attraction is central, and the magnitude of the central attraction is (§ 43) $\ddot{r} - r\dot{\theta}^2$. Now the

equation of any conic section referred to a focus as pole is $r = 1/a\ (1 + e \cos \theta)$, a being the semi-axis major, and e being the eccentricity. Hence (writing $r^2\dot{\theta} = h$, from which we at once find

$$\ddot{\theta} = -\frac{2}{r}\frac{dr}{d\theta}\dot{\theta}^2\Big)$$

we easily obtain by the methods of Chap. V., $\ddot{r} - r\dot{\theta}^2 = -ah^2/r^2$, which shows that the force is attractive and varies inversely as the square of the distance. Finally, since $r^2\dot{\theta} = h$, we have $a^2(1 + e \cos \theta)d\theta = hdt$; and the integral of this throughout a complete revolution, *i.e.*, from $\theta = o$ to $\theta = 2\pi$, is $ht = 2\pi a^2$. But if μ be the acceleration at the mean distance, a, so that $h^2 = \mu a$, we find $4\pi^2a^3 = \mu^2t^2$. And so, from the third law of Kepler, we deduce the result that gravity depends only on the quantity, and not on the quality, of matter; for any two bodies, having the same mean distance from the sun, would have the same periodic time, provided only that their masses were the same.

87. If, from the above evidence, we now assert Newton's great Law of Gravitation that *Every particle of matter in the universe attracts every other particle with a force whose direction is that of the line joining the two, and whose magnitude is directly as the product of their masses, and inversely as the square of their distance from each other*, we see that Kepler's Laws cannot be strictly true. In the first place all the bodies in the solar system (including the sun) will revolve about the centre of inertia of the system, which is not necessarily nor actually situated at the sun's centre; and again, their paths cannot be true ellipses because of mutual attraction.

But we *know* that Kepler's Laws are not strictly true; and, further, the deviations from them are precisely such as should result from mutual attraction amongst the various planets. Indeed, by assuming the truth of Newton's Laws of Motion and of the Law of Gravitation, Adams and Leverrier were able to predict independently the existence and position of the previously unknown planet Uranus.

The law of gravitation is supported by almost as strong proof as any theoretical statement can possibly have.

88. So far as we have gone we have looked upon the sun and the planets as mere material points. The justification of this is contained in the latter of two theorems due to Newton: 1°, *A spherical shell composed of uniform gravitating matter exerts no resultant attraction upon a particle in its interior; and*, 2°, *it*

attracts an external particle as if its whole mass were condensed at its centre.

Let A be a point external to the spherical shell PQQ'P', the centre of which is at O, and the density of which per unit surface is ρ. Draw a cone, APQ, of infinitely small angle ω, having its vertex at A, and intercepting small surfaces of the sphere at P and Q. These surfaces are equally inclined to APQ, and so the masses which are cut off at P and Q, being equal to $\omega\rho AP^2$ sec OPQ and $\omega\rho AQ^2$ sec OPQ respectively, attract a particle at A equally. Now take an exactly similar cone at AP'Q', equally inclined to, but on

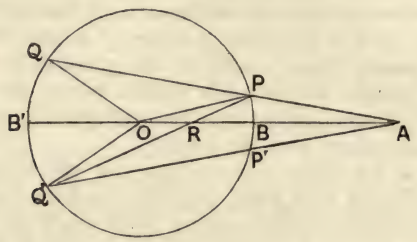


FIG. 56.

the opposite side of, AO. The elements of the shell at P and Q attract the particle at A equally. Let PQ' intersect AO in R. The position of R is obviously independent of that of P, and hence R is the vertex of a cone, of angle ω', say, which intercepts the same elementary areas of the shell at P and Q' as the cones APQ and AP'Q' intercept. The masses are therefore $\omega'\rho PR^2 \sec OPR$ and $\omega'\rho Q'R^2 \sec OPR$ respectively; and, since OPR = OAP, their resultant attraction on unit mass at A is

$$\omega'\rho\left(\frac{PR^2}{AP^2}+\frac{Q'R^2}{AQ'^2}\right).$$

But PR/PA = BR/BA = OP/OA = Q'R/Q'A, and so the resultant is

$$2\omega'\rho\,\frac{OP^2}{OA^2}.$$

Hence, summing all such quantities for each similar pair of elements, the whole attraction of the shell is $4\pi\rho$ OP^2/OA^2 in the direction AO. And so the proposition is proved, since $4\pi OP^2$ is the whole surface of the shell.

Since the proposition is true of a shell, it is also true of a solid

sphere, which may be regarded as built up of a number of such shells. It is, therefore, practically true of the planets and of the sun; for the planets, even although in no case strictly composed of uniformly dense concentric layers, are yet at distances from the sun which are large in comparison with their own dimensions.

[When A is inside the shell, P and Q are on opposite sides of it, and so the truth of the first theorem above is established.]

The above case furnishes one example of that limited class of bodies which attract, and are attracted by, external bodies, as if their whole mass were condensed at a definite point, called their

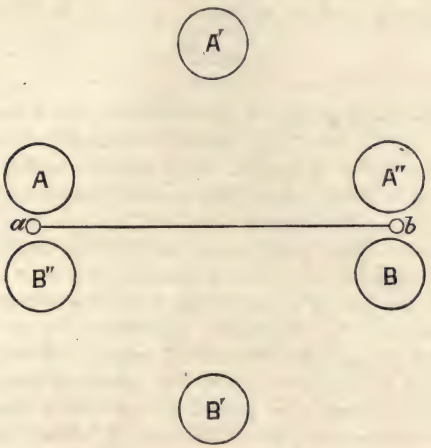


FIG. 57.

centre of gravity. Another example is given in § 317. The centre of gravity, when it exists, always coincides with the centre of inertia.

89. The second theorem is made use of in Cavendish's method of determining the mass (and, consequently, the mean density) of the earth.

Two small leaden balls, of mass m, are attached to a light rigid rod or tube, ab, which is attached at its middle point to a vertical wire. The couple required to twist the wire through a given angle is determined by observations upon the time of oscillation of the system. Two large leaden balls of mass M, which were originally in the positions A′ B′, are placed in the positions A, B. The mutual attractions of the balls A and a, B and b, deflect ab from its

normal position, and it oscillates about a new position. The angle through which ab is deflected is determined by means of a beam of light reflected from a mirror which is fastened to the suspending wire. Similar observations are made with the large balls once more in the positions A′, B′, and finally in the positions A″, B″, at the same distance as before from ab, but on opposite sides of them, so as to deflect ab in the opposite sense. The mean of the deflections is taken, and the couple required to produce it is known. In this way the attraction between the two masses M and m, their centres being at a given distance r apart, is found; and it may be compared with the attraction of the earth upon the small ball. The mass of the earth being μ, and its radius R, we have $k\mu/R^2 = M/r^2$. This gives

$$k\rho\tfrac{4}{3}\pi R = \frac{M}{r^2},$$

where ρ is the mean density of the earth, and k is a constant.

Professor C. V. Boys has recently succeeded in drawing out extremely fine fibres of quartz, the torsional rigidity of which is so small that, by their means, the mutual attraction between two lead pellets can easily be made manifest.

90. The *Schehallien* experiment, undertaken with the object of determining the value of ρ, was of precisely the same nature. The deflection of the bob of a pendulum from the true vertical under the attraction of the mountain Schehallien was obtained by determining the apparent difference of latitude of two places—one being situated to the north, and the other to the south of the mountain—by means of a pendulum used as a plumb-line. If from this we subtract the true difference of latitude of the places, we get a measure of the attraction of the mountain upon the bob of the pendulum as compared with the attraction of the earth upon it; for the attraction of the mountain pulls the pendulum from the vertical in such a way as to increase the apparent latitude of the northern point, and to diminish that of the southern. The great objection to this method lies in the fact that our knowledge of the mass of the mountain is necessarily very imperfect.

Yet another method consists in observing the times of oscillation of pendulums of the same length, one of which is situated at the top, and the other at the bottom, of the shaft of a coal-pit. In this way we get a comparison of the attractions of the whole earth, and of the earth minus a shell of thickness equal to the depth of the mine, upon a mass situated at known distances from the centre. Here, again, uncertainty arises, because of our limited knowledge of the density of the earth's crust, which must be assumed to be

equal to that in the neighbourhood of the shaft. This is called the *Harton* experiment, because it was performed at the Harton mines.

The value of the mean density of the earth obtained by the Schehallien experiment is considerably less than, and the value obtained by the Harton method is considerably greater than, the mean of the (very closely accordant) results obtained by different experimenters who used the Cavendish method. Their results show that the mean density is about 5·5 times the density of water.

Hypotheses Framed to Explain Gravitation.

91. Various attempts at an explanation of gravitation have been made—all more or less unsatisfactory.

One of the most noted of these is Le Sage's hypothesis of ultra-mundane corpuscles. According to Le Sage, gravitation is due to the bombardment of bodies by numberless small material particles which are darting about in space with great speed in all directions. A single material body placed in space would not be impelled in any one direction more than in another, for the corpuscles would batter it equally on all sides. But two bodies in space would be driven together—provided, at least, that their distance apart were small in comparison with the free-path (§ 148) of the corpuscles—for those sides of the bodies which face each other would be shielded to a greater or less extent from the bombardment. If the dimensions of the bodies were very much smaller than their distance apart, the force of attraction (the incongruity of which term is evident from the fact that the force is here really one of *impulsion*) would vary in direct proportion to the cross-sectional area of each of the two bodies as seen from the other. But this is not the gravitational law. In order to obtain it, we must make the supposition that the molecules of matter are so far apart that the number of corpuscles which pass completely through (say) the earth, without striking any part of it, is enormously larger than the number of those which are stopped by it. In this case, every molecule in the interior will be bombarded equally with an exterior particle. The force also will be inversely as the square of the distance, and so we get Newton's Law.

In order that the planets should not experience appreciable resistance to their motion around the sun, it is necessary to assume, further, that the speed of the planets is zero relatively to that of the corpuscles. And this indicates one great defect of the hypothesis; for the energy of the corpuscles which must be spent

in the maintenance of gravitation would be sufficient, by its transformation into heat, to completely volatilise any material substance of which we have knowledge.

92. The law of gravitation may be worked out into all its consequences (at least, so far as our methods avail) without any knowledge of the mechanism by which it occurs. We require merely to assume that it acts directly at a distance. But, as Newton remarked, no one competent to think correctly on physical matters will be content with this assumption. He himself suggested the rarefaction of the ether in the neighbourhood of dense bodies as a possible explanation. Sir W. Thomson has pointed out a dynamical method of producing this diminution of pressure. An incompressible fluid, filling all space, which is brought into existence, or is annihilated, at the surface of every particle of matter at a rate proportioned to the mass of that particle, and which is annihilated, or produced, at an infinite distance, at the same total rate would supply the necessary means. (See § 74.)

Waves traversing a medium would have the effect of making bodies immersed in it approach each other.

The property of *dilatancy* (Chap. XXXIII.) in a medium composed of rigid particles in mutual contact would also account for a gravitational action on bodies placed in it.

A certain stress in Maxwell's electro-magnetic medium (Chap. XXXII.) would account for it too. So, as we have seen, would Le Sage's ultra-mundane corpuscles. But, to every provisional hypothesis yet brought forward, some objection, more or less conclusive, may be advanced.

If gravitation be due to action propagated through a material medium, its propagation through finite distances must occupy a finite time. We can assert merely that the speed of propagation is large in comparison with planetary velocities; for no such modification of the planetary motions, as would be entailed were it otherwise, is observable.

The Nebular Hypothesis.

93. Various nebulæ, when examined under sufficiently high magnifying power, are seen to be merely groups of stars like our own stellar system. But others cannot be so resolved; and, among them, there is great variety of constitution. Some appear to be comparatively uniform throughout, while others seem to vary greatly in density at different parts. Others, again, have very dense nuclei with a faint nebulous surrounding.

Such facts as these suggested to Laplace his famous Nebular

Hypothesis. According to this hypothesis, we see in these nebulæ solar systems, such as our own, in various stages of formation. The sun was formed by the mutual gravitation of its parts from a state of diffusion throughout space. And, to account for its rotation, these parts must have had the same amount of moment of momentum about an axis as that which the sun at present has. In all probability they had small translational velocity before they began to gravitate towards each other—small, that is, relatively to the speed which they would acquire under the action of gravitation; for, if this were not so the chance of their colliding would be vanishingly small. The collision of these parts would produce great heat, which would result in the production of a nebula, extending beyond the orbit of Neptune, and slowly revolving on its axis. As the nebula gradually shrank, its angular velocity would increase until, by 'centrifugal force,' a ring of matter was left behind as the body of the nebula still farther shrank. The breaking up of this ring (which would usually occur, as dynamical principles show), and the subsequent agglomeration of its parts, would result in the formation of a planet. And so the development would proceed.

Laplace's hypothesis cannot, as he originally stated it, give a *full* account of all the phenomena of our solar system; but this we do know, that some such hypothesis must be true if our sun's heat has had a physical origin such as our present knowledge is adequate to explain. We cannot conceive of any store of energy, sufficient to account for the immense radiation of heat from the sun, except the potential energy of separated masses of gravitating matter.

Potential.

94. Since any force f is equal to the space-rate of variation of the kinetic energy of the material system upon which it acts, and since, by the principle of conservation, the change of kinetic energy is equal and opposite to the change of potential energy, we may write

$$f = \frac{de}{ds} = -\frac{dV}{ds},$$

where V represents the potential energy.

The value of the quantity V depends upon the instantaneous position of the material system, and also upon the mutual configuration of its various parts.

We may define the mutual potential energy of two material bodies in any given relative position, as *the amount of work which may be obtained by allowing them to move, under their mutual repulsion, to an infinite distance apart.* And, of course, this

definition implies that we choose the configuration of infinite distance as the configuration of zero potential energy.

The *Potential* at any point, due to any given distribution of matter, is the mutual potential energy between that matter and a unit of matter placed at the given point. To find its value we have only to multiply the force at any point of the path, along which the unit of matter is repelled, by the infinitesimal element, ds, of the path, and to sum all such quantities from the given position to an infinite distance. This quantity is represented by the symbol

$$\int_s^\infty f ds,$$

the meaning of which is that we are to find the general value of the integral of fds; to replace in it, first, s by ∞; second, s by its actual value at the given point; and, finally, to subtract the latter quantity, so found, from the former. This makes the work which is done depend only upon the initial and the final positions of the unit of matter—a condition which must be satisfied if the forces are consistent with the principle of conservation. For, if the work depended upon the path along which the matter was repelled, we might cause it to return, by frictionless constraint, to its former position, by a path which necessitates an expenditure of less work than that which was done upon it by repelling forces. In such a case there would consequently be a continual gain of energy without any corresponding expenditure.

Let us suppose, for example, that the mutual force is repulsive and inversely proportional to the square of the distance, s, between the two portions of matter. Let s increase from s to s'. The work which is done is

$$\int_s^{s'} \frac{m}{s^2} ds$$

where m is the mass of the repelling matter, and we assume that the law of repulsion is similar to Newton's law of attraction, and define unit force as the force between two unit masses placed at unit distance apart. The general value of the integral is (§ 35, example (3))

$$-\frac{m}{s}$$

and the work is therefore

$$\frac{m}{s} - \frac{m}{s'}.$$

If we now put $s' = \infty$, the second term disappears and we see that the value of the potential is

$$V = \frac{m}{s},$$

and the repulsive force is

$$\frac{m}{s^2} = -\frac{dV}{ds}.$$

95. *Gravitational Potential.*—When the force is attractive, as in the case of gravitation, the potential—defined as above—becomes

$$V = -\frac{m}{s}.$$

The potential is therefore essentially negative; that is to say, work is done *against* the forces when the distance between the particles is increased. But it must carefully be observed that the fact of the potential energy being negative, at distances less than infinite, is due entirely to our having, for convenience, chosen infinite distance apart as the configuration of zero potential. The increase of potential energy is positive as we pass towards infinity. Still, to avoid the inconvenience of defining the gravitational potential as a negative quantity at all finite distances, it is preferable to change the sign and write, in this case also,

$$V = \frac{m}{s}$$

$$\text{and } f = -\frac{dV}{ds},$$

where f is to be understood as signifying *inward* or *attractive* force. The potential V, therefore, does not represent the mutual potential energy. It is the *exhaustion* of potential energy which occurs when the unit of matter passes from infinity, under the attractive forces, to the position s.

The potential at any point, due to a number of separate masses, is simply the sum of the separate potentials due to each mass.

96. *Equipotential Surfaces. Lines and Tubes of Force.*—Any surface over which V is constant is called an Equipotential Surface; and a line, the tangent to which at any point is always in the direction of the force at that point, is called a *Line of Force.* Each line of force is a possible path along which a material particle would move under the given forces.

Since dV is zero as we pass from one point to another point of an equipotential surface, it follows that the force at any point of such a surface has no component along it. In other words, the lines of force are everywhere perpendicular to the equipotential surfaces.

No two different equipotential surfaces can intersect with one another; for this would imply that a finite change of potential would follow an infinitely small displacement of a material particle, *i.e.*, the force would be infinite at any such intersection, and no examples of infinite forces occur in nature.

If, through every point of an infinitely small closed curve drawn on an equipotential surface, we draw lines of force, a tube of infinitely small section will be formed. Such a tube is called a *Tube of Force.*

Necessarily, the lines and tubes of force can originate only at a point where matter is situated; for the force owes its origin to the presence of matter. And if, at any point of space, a force f exists, we may draw an infinitesimal tube of force, so as to contain that point and to enclose f lines of force per unit area of its normal section. The number of the lines of force per unit of sectional area therefore indicate the intensity of the force at the given point.

If a given portion of a tube does not contain any matter, the number of lines which it contains remains constant, since no line can end in its interior, and none can pass out through its sides. This gives

$$f\sigma = c,$$

where f is the force at any point of the tube at which the sectional area is σ, and c is a constant, numerically equal to the number of lines which the tube contains.

97. *Special Applications.*—(1) Let the attracting body be symmetrical about a point. In this case the tubes of force are cones, and the area of any normal section of each of them is proportional to the square of its distance from the point of symmetry. Hence the above equation shows that the force at any point is inversely proportional to the square of the distance of that point from the point of symmetry, say

$$f = \frac{a}{s^2}.$$

(2) Let the attracting matter be symmetrically arranged about an infinitely long straight axis. The equipotential surfaces are concentric cylinders whose common axis coincides with the axis of symmetry, and the tubes of force are wedges bounded by axial planes. The section is proportional to the distance from the axis, and therefore the force is inversely proportional to that distance, *i.e.*,

$$f = \frac{b}{s}.$$

(3) Let the matter be arranged homogeneously in infinite parallel planes. The equipotential surfaces are planes parallel to these, and the tubes of force are cylinders arranged perpendicularly to the planes. Consequently the force is constant at all distances, say

$$f = c.$$

98. *Total Force over a Closed Surface.*—Draw any closed surface S (Fig. 58), and,

(1) Let m be a massive particle, outside the surface, to which the force is due. Draw any infinitesimal tube of force, mnp, cutting the surface at n and p. (Of course, it may cut the surface in any *even* number of places.) The total force over the portion of the surface intercepted by the tube at n, is equal to that over the portion intercepted at p; but, in the one case, the force is directed outwards over the surface, while, in the other, it is directed inwards. Hence, by consideration of an infinite number of such

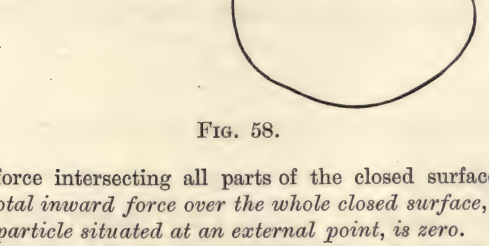


FIG. 58.

tubes of force intersecting all parts of the closed surface, we see that the *total inward force over the whole closed surface, due to a material particle situated at an external point, is zero.* [The use of the word *inward*, of course, implies that the force is to be reckoned as negative when it is outwardly directed.]

The same proof applies in the case of any number of material particles.

(2) Let the particle, of mass m, be placed within the closed surface. Draw a sphere, with unit radius, from the point m as centre. The area of this spherical surface is 4π, and the force at any point of it, due to the attraction of the central particle, is equal to m. The total inward force is therefore $4\pi m$. But, by the result of § 96, the total inward force over this surface is equal to that over the given surface S. Hence, if m be the whole amount of matter contained within S, *the total inward force over the whole closed surface, due to matter of amount* m *enclosed within it,* is $4\pi m$.

These results may be symbolised thus,

$$\int N dS = 0 \text{ or } 4\pi m,$$

where N rep esents the number of lines of force which cross the surface S per unit of area in the part where the element of surface dS is taken.

99. *Special Applications.*—We may apply these results to the determination of the values of the constants a, b, and c, in § 97.

In example (1) of that section, the whole number of lines of force which cross any closed surface surrounding the spherical distribution of attracting matter is $4\pi m$, where m is the whole amount of matter. Hence, if we suppose the enclosing surface to be a concentric sphere of radius s, we get as the total force

$$4\pi s^2 f = 4\pi s^2 \frac{a}{s^2} = 4\pi m.$$

Therefore

$$a = m.$$

Similarly, if, in example (2), we consider the force due to the matter m contained in unit of length of the cylindrical distribution, we get

$$2\pi s f = 2\pi s \frac{b}{s} = 4\pi m,$$

for $2\pi s$ is the area of unit of length of a concentric cylinder of radius s.

Therefore $$b = 2m.$$

Again, if (example (3)) we draw a right circular cylinder of unit radius, perpendicular to the infinite planes, and close its ends by parallel planes on opposite sides of the given plane distribution of matter, the force exerted over the two ends is

$$2\pi f = 2\pi c = 4\pi m,$$

so that $$c = 2m.$$

In particular, if the given distribution consists of an infinitely thin plane layer, of infinite extent, and of finite surface density σ, the force at any point is

$$f = 2m = 2\pi\sigma.$$

[Such a distribution cannot occur in the case of gravitational matter, but the problem has a direct application in the theory of electricity.] It follows that the force at any point *just outside* a surface on which matter is distributed is normal to the surface and

equal to $2\pi\sigma$—provided that the surface is of finite curvature; for the point may be taken so close to the surface that, as seen from it, the surface is practically an infinite plane; that is to say, any infinitely small portion of the surface is plane, and the point may be taken infinitely close to this portion in comparison with its dimensions, infinitely small though they be.

100. In the case just considered, the normal force at the two sides of the surface are respectively $f = 2\pi\sigma$, and $f' = -2\pi\sigma$. The total change of force in crossing the surface is therefore

$$f - f' = 4\pi\sigma.$$

This shows us how *to distribute matter over a given surface in order to produce a given change in the value of the normal force in passing from one side to the other.*

We may write this expression in the form

$$-\left(\frac{dV}{dn} - \frac{dV'}{dn'}\right) = 4\pi\sigma,$$

where V and V′ are the values of the potential at each side of the surface, while dn and dn' are measured along the outwardly drawn normals on each side. This enables us to calculate the distribution of matter when the discontinuous distribution of potential is given.

It is easy also to obtain an expression for the volume-density of matter which is required to produce a given continuous distribution of potential.

For, since the force f outside a symmetrical spherical distribution of matter is (§ 99) given by the equation

$$4\pi r^2 f = 4\pi m,$$

the value of f just outside the sphere is $f = m/r^2$ (which, we may observe in passing, proves Newton's theorem, § 88). But m is equal to $4/3 \cdot \pi\rho r^3$, if the sphere is of uniform density ρ; in which case, therefore, $f = 4/3 \cdot \pi\rho r$. That is

$$-\frac{dV}{dr} = \frac{4}{3}\pi\rho r,$$

whence

$$-V = \frac{2}{3}\pi\rho r^2 + C \text{ (a constant)}.$$

Now, if we take the origin of co-ordinates at the centre of the sphere, we have

$$r^2 = x^2 + y^2 + z^2;$$

whence $dr/dx = x/r$, $dr/dy = y/r$, $dr/dz = z/r$, for, since x, y, and z,

are independent variables (§ 28), we must assume y and z to be constant when x varies, and so on. Therefore we get

$$-\frac{dV}{dx}=\frac{4}{3}\pi\rho r\frac{dr}{dx}=\frac{4}{3}\pi\rho x$$

$$-\frac{d^2V}{dx^2}=\frac{4}{3}\pi\rho.$$

And, similarly, $$-\frac{d^2V}{dy^2}=\frac{4}{3}\pi\rho, \quad -\frac{d^2V}{dz^2}=\frac{4}{3}\pi\rho,$$

whence $$-\left(\frac{d^2V}{dx^2}+\frac{d^2V}{dy^2}+\frac{d^2V}{dz^2}\right)=4\pi\rho.$$

Around any point in space throughout which matter is distributed with density ρ, describe a sphere which is so small that the density of the matter which it contains is sensibly constant. We may suppose all the attracting matter to consist of two portions—that which is within the little sphere, and that which is external to it. We may divide the whole potential V into two parts, V_1 and V_2, of which the former is due to the sphere and the latter is due to the matter external to the sphere. V_1 therefore satisfies the equation

$$-\left(\frac{d^2V_1}{dx^2}+\frac{d^2V_1}{dy^2}+\frac{d^2V_1}{dz^2}\right)=4\pi\rho,$$

ρ being the density of the matter which produces the potential V_1 at the given point. But the matter outside the sphere does not contribute to the density at the given point within. That is, at the given point, the density of the matter which produces the potential V_2 is zero, and therefore

$$-\left(\frac{d^2V_2}{dx^2}+\frac{d^2V_2}{dy^2}+\frac{d^2V_2}{dz^2}\right)=0.$$

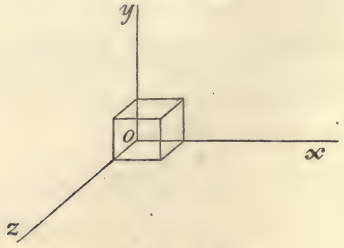


FIG. 59.

Consequently, by addition, we get quite generally

$$-\left(\frac{d^2V}{dx^2}+\frac{d^2V}{dy^2}+\frac{d^2V}{dy^2}\right)=4\pi\rho$$

This shows us how *to distribute matter throughout space, so as to produce a given continuous distribution of potential.*

The above result may be obtained in a totally different manner, which is even simpler, and which will make the meaning of the equation more evident.

Take three rectangular axes at any point o (Fig. 59), and draw a little parallelepiped at this point with its edges parallel to the x, y, and z axes; and let the lengths of the edges be δx, δy, δz, respectively.

Resolve the forces in the neighbourhood of the parallelepiped into their components parallel to the axes. This will not alter the result of § 98; and we may now draw the lines of force perpendicular to the various small surfaces.

Let n_x be the number of lines of force which cross unit area of the face of the parallelepiped which passes through the origin and is perpendicular to the x axis. The total number of lines which cross that face is therefore $n_x \delta y \delta z$, since $\delta y \delta z$ is the area of the face. Similarly the number which cross the parallel face at the distance δx from the former is

$$\left(n_x + \frac{dn_x}{dx}\delta x\right)\delta y \delta z.$$

Now the lines which cross the former face are due entirely to matter outside the little volume, and they therefore cross the parallel face also. Hence the total number of lines which enter the little volume from without by the two faces is

$$\frac{dn_x}{dx}\delta x \delta y \delta z,$$

the difference of these two quantities. By similar reasoning for the other pairs of faces we find that the total number of lines which enter the little volume (that is, the excess of those which enter over those which leave) is

$$\left(\frac{dn_x}{dx} + \frac{dn_y}{dy} + \frac{dn_z}{dz}\right)\delta x \delta y \delta z.$$

But, by § 98, this is equal to $4\pi\rho\delta x \delta y \delta z$, where ρ is the density of the matter contained in the volume $\delta x \delta y \delta z$. Therefore

$$\frac{dn_x}{dx} + \frac{dn_y}{dy} + \frac{dn_z}{dz} = 4\pi\rho.$$

And the result is independent of the size of the little volume; so that the meaning of the equation is that the volume density at any point of space is $1/4\pi$ times the number of lines of force which originate, per unit of volume, at that point.

The following proposition is also of great importance. *It is possible so to distribute matter over a given surface, which encloses a given mass, as to produce, outside that surface, the same potential as that which the given mass produces.* The mathematical proof of this proposition cannot be introduced here, but an experimental proof will be given in the chapter on electrostatics.

101. The calculation of the distribution of potential which is produced by a given distribution of matter is extremely difficult or even impossible in most cases; but the beautiful method of *electric images*, due to Sir W. Thomson, enables us to deduce with great ease the solution of many unknown problems from the known solution of others. The further discussion of this subject may be left until we treat of the subject of electricity.

CHAPTER IX.

PROPERTIES OF GASES.

102. *Compressibility.*—Throughout this investigation we assume that the temperature of the gas remains constant. The effects which result from changes of temperature will be more conveniently treated in the chapter on the effects of heat.

All gases are compressible; that is, their volume can be diminished by the application of pressure. We shall see afterwards that sound could not pass at a finite rate through a gas which was not compressible. So that the mere fact that gases can convey sound constitutes a proof of their compressibility.

103. *Boyle's Law.*—The law which very completely, though not with absolute accuracy, represents the relation between the pressure and the volume of air (and many other gases) was discovered experimentally by Boyle, who showed that *the density of a gas is directly proportional to the pressure.* In symbols, ρ being the density and p the pressure, this is

$$\rho = cp,$$

c being a constant. The density, that is the mass or quantity of matter in unit volume, is numerically equal to the reciprocal of the volume containing unit mass. Hence, v being this volume, we may write, instead of the above,

$$pv = c,$$

where the constant c is the reciprocal of the former one; or, in words, *the volume is inversely proportional to the pressure.*

Boyle's apparatus consisted of a glass **U**-tube (Fig. 60) with a long and a short limb. The long limb was open to the atmosphere, while the short one was closed, and contained a quantity of air, which was separated, by means of mercury filling the bend of the tube, from the outside air. The level of the mercury was the same in both limbs of the tube, and so the enclosed air (the volume of which was carefully noted) was at atmospheric pressure (§ 75). Mercury was then poured into the open limb, until a difference of level equal to

the height of the mercury barometer was established. The air inside was therefore under a pressure of two atmospheres, and its volume was found to have been halved; and so on with other values of the pressure.

A slight modification of the apparatus enables us to prove the law under diminished pressure. AB (Fig. 61) is a vessel containing mercury. The glass tube *ab*, which is closed at the end *a*, but is open at *b*, is filled with mercury, and inverted in AB. Being shorter than the height of the mercury barometer, the tube *ab* remains filled. Air or any other gas may now be introduced into it until the mercury inside is at the same level as that outside. Under these conditions the gas in *ab* is under atmospheric pressure. If *ab* be raised, the mercury in it stands at a higher level than that outside, and the gas expands, since it is under diminished pressure. The ideal gas which rigidly obeys Boyle's Law is called a *perfect gas*.

FIG. 60.

FIG. 61.

104. *Compressibility of a Perfect Gas.*—One gas is more compressible than another in direct proportion to the alteration of volume produced by a given pressure, and in inverse proportion to the pressure required to change the volume to a given extent. Hence we measure the compressibility by the ratio of the percentage change of volume to the change of pressure which produces it. That is to say, if the volume V changes by the quantity v when the pressure alters by the amount p, the compressibility is measured by the ratio $v/\mathrm{V}p$.

By Boyle's Law we have

$$\mathrm{PV} = c,$$

and

$$(\mathrm{P}+p)\,(\mathrm{V}-v) = c.$$

Therefore

$$p\mathrm{V} - \mathrm{P}v = 0,$$

since we can neglect pv, which is the product of two small quantities. This gives

$$\frac{v}{\mathrm{V}p} = \frac{1}{\mathrm{P}},$$

that is, *the compressibility of a perfect gas is inversely proportional to the pressure.*

105. *Deviations from Boyle's Law.*—Though no gas is perfect, yet many gases do not greatly deviate from Boyle's Law throughout a considerable range of pressure.

Air is more compressed than it should be in accordance with the

law until a pressure of nearly one ton's weight on the square inch is reached. After this point, the compression is *less* than the calculated value. A reason for this is simply that the volume of the gas is not capable of indefinite decrease, while Boyle's Law asserts that under infinite pressure the volume will become zero.

Hydrogen, unlike air, is, at ordinary temperatures, always less compressible than the law indicates. Nitrogen, along with many other gases, resembles air.

These results are exhibited graphically in Fig. 62. The actual

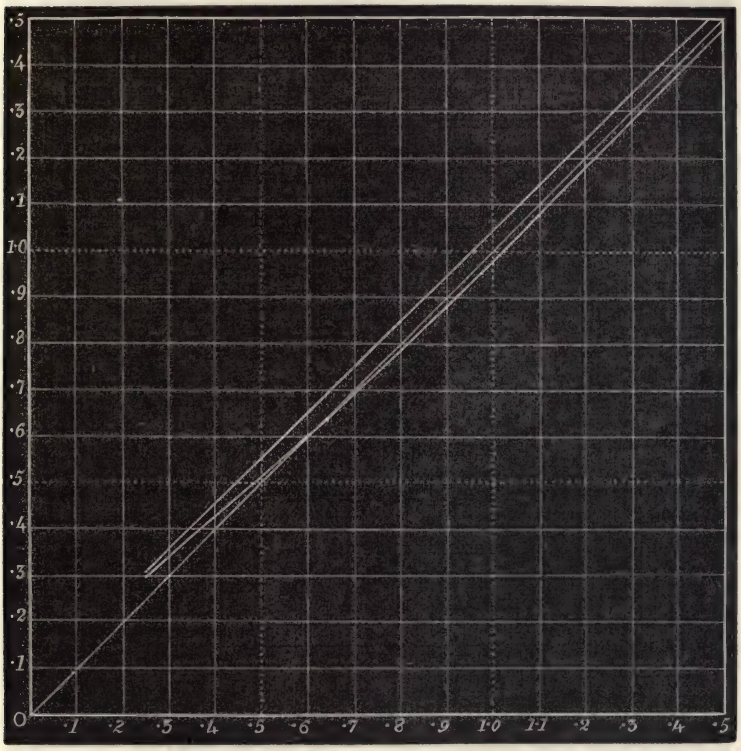


FIG. 62.

volume of the gas is measured along the vertical axis, while the volume of a perfect gas under the same pressure is measured along the horizontal axis. The straight line passing upwards through the origin at an angle of 45° is obviously the graph for a perfect

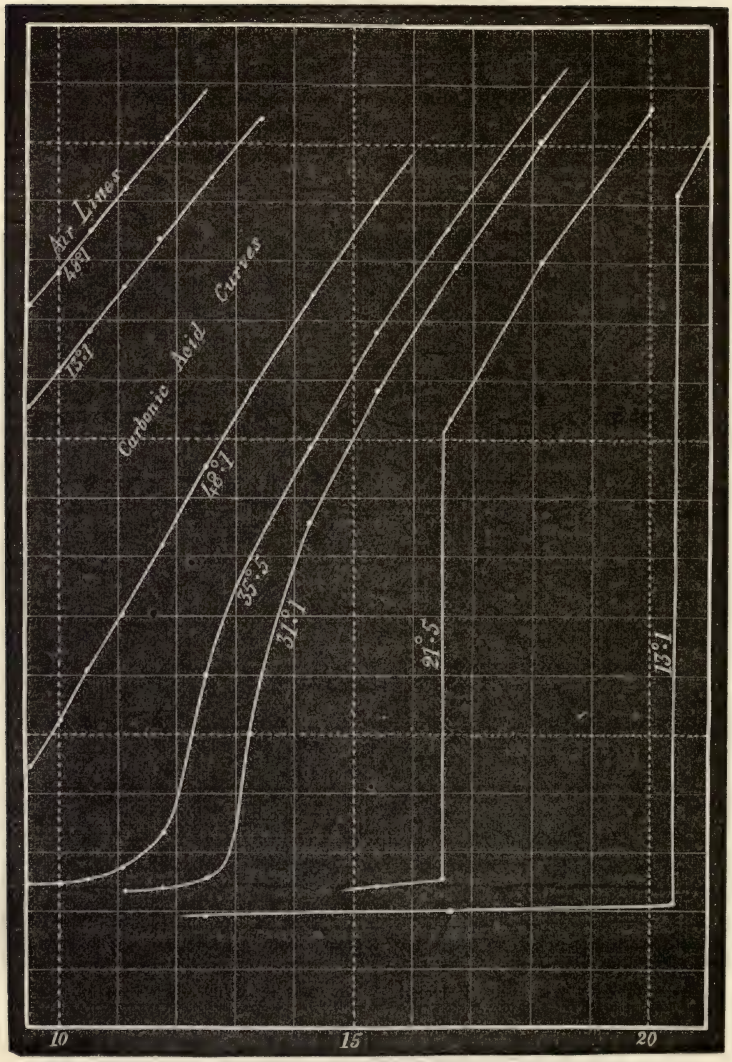

Air Lines
48°1
13°1
Carbonic Acid Curves
48°1
35°.5
31°.1
21°.5
13°.1
10
15
20

FIG. 63.

gas. The curved line intersecting the perfect gas-line, and, like it, sloping upwards towards the right, represents the action of air; and the other curve, sloping similarly, is the graph for hydrogen.

106. *Compression of Vapours.*—A vapour, though it may obey Boyle's Law throughout a considerable range of pressure, ultimately deviates more and more from that law as the pressure rises. The direction of the deviation is similar to that of air at pressures less than 152 atmospheres, *i.e.*, the vapour is more compressed than a perfect gas would be. When the pressure has become sufficiently great, the vapour begins to liquefy; and the pressure then remains constant until the whole has become liquid. Further compression is comparatively a matter of extreme difficulty.

The whole process above described must take place with extreme slowness in order that the condition (§ 102) of constant temperature may be adhered to.

We may now suppose the temperature to be increased to, and maintained at, a definite value higher than that which it formerly had. If the pressure has the same value as it had at the commencement of the former process, the volume of the vapour will be greater than before, for all vapours expand when heated under constant pressure. And, if the pressure be increased as in the previous case, a precisely similar series of phenomena will be presented; the volume of the substance, however, being always larger than formerly under the same conditions as regards pressure. But one important difference will be noted—the change of volume during the process of liquefaction will be less than it was when the temperature was lower. Ultimately, when the temperature is sufficiently high, there is no sudden change of volume when the substance assumes the liquid condition. At still higher temperatures, the deviations from Boyle's Law become less and less marked.

At all temperatures above the limiting one at which sudden liquefaction ceases, the substance is called a *gas;* at lower temperatures it is termed a *vapour.*

With this explanation, no difficulty will be experienced in understanding the diagram on the opposite page. The volume, v, of a perfect gas is measured along the horizontal axis from a point not shown in the diagram. The scale is such that 1,000 times the reciprocal of the abscissa represents the pressure in atmospheres. The actual volume, v', of carbonic acid gas is measured along the vertical axis. In this way a series of curves are shown which indicate the deviation of that gas from Boyle's Law at various temperatures. Portions of two of the nearly straight lines, which these curves would become if the gas were air, are drawn. The vertical portions of two of the

curves indicate the stage during which liquefaction occurs—the almost horizontal parts belong to the liquid carbonic acid.

The substance is more or less compressible under given conditions of pressure and temperature than a perfect gas is, according as a line touching the curve (at the point satisfying these conditions) makes a greater or less angle with the horizontal axis than the angle whose tangent is v'/v times the tangent of the angle made by the corresponding perfect-gas line. Hence the substance, when in a condition resembling the liquid state, is less compressible than a perfect gas would be; and we thus see that hydrogen is less compressible than a perfect gas, because, under ordinary conditions of temperature and pressure, it is in a state more analogous to that of liquids than to that of gases. Were it examined under conditions similar to those of carbonic acid in the upper right-hand region of the diagram opposite, *i.e.*, under sufficiently diminished temperature and pressure, it would almost certainly be found to have a smaller volume than Boyle's Law shows.

107. *Elasticity.*—All gases and vapours possess *perfect* elasticity of bulk. That is to say, they entirely recover their original bulk when allowed to do so by means of the removal of the distorting pressure. This may readily be proved by the simple experiment of inverting in water a glass vessel containing air or any gas which is not appreciably dissolved by the liquid. The gas may be subjected to, or relieved from, pressure by raising or lowering the glass vessel. The possibility of discharging a bullet from a gun, or of propelling a vessel or driving machinery by means of compressed gases furnishes another proof of their elasticity. And still another proof consists in the fact that they all convey sound, which would be impossible were they not elastic, just as it would be impossible if they were not capable of being compressed.

108. *Viscosity.*—We have already given a general definition of this term as the property in virtue of which there is resistance to shearing motion. But it is convenient to use the word as referring to a specific property (one independent of the size of the body, § 81). Hence we define viscosity as the tangential force per unit area of two indefinitely large parallel plane surfaces of the fluid which are at unit distance apart and move parallel to each other with unit relative speed. It follows that the tangential force per unit area of two such planes at a distance x apart, and moving with relative speed v, is $\tau v/x$, where τ is the viscosity. But, in shearing motion v is always proportional to x, so that the tangential force is $\tau dv/dx$.

In making an actual determination of the value of τ in any gas,

various forms of experiment based upon the above definition might be used. Clerk-Maxwell used a circular disc which vibrated torsionally about a perpendicular axis through its centre. Two similar *fixed* discs were placed one on each side of the vibrating disc, and the gas occupied the intervening space. The disc would obviously oscillate more slowly in a viscous gas than in one which possessed small viscosity; and the quantity τ may be determined from the results of such experiments. The mathematical investigation is somewhat more difficult than we can venture to introduce into an elementary work.

This property varies very much from one gas to another. In hydrogen, carbonic acid, air, and oxygen, it increases from the first-mentioned to the last-mentioned, being about half as great in hydrogen as in air.

It *increases* very markedly also with rise of temperature.

The slow descent of clouds, or of fine suspended dust, in air is due to the viscosity of that gas. The weight of a drop of water, which causes its descent, is proportional to the cube of its diameter; but the resistance which results from viscosity is proportional only to the first power of the diameter. Hence, if the diameter of a drop be reduced to one-tenth of its original value, the weight becomes one-thousandth of what it was before, while the resistance is merely reduced to one-tenth of its previous amount. That is to say, the resistance is relatively one hundred times more effective than formerly.

109. *Diffusion.* — When two gases, which are not intimately mixed, occupy a certain volume, each gradually diffuses itself throughout the whole volume, so as to fill it just as it would have done had the other been absent. The only effect of the presence of the other gas (on the presumption that there is nothing of the nature of chemical action between them) is that the time taken by the first to uniformly fill the space is greatly increased. The above process is called *diffusion*, and the corresponding property is *diffusivity*.

Experiment shows that the quantity of gas which passes in time t through an area a, perpendicular to which the rate of variation of density per unit of length is r, is proportional conjointly to t, a, and r. Hence, if q be this quantity, we have

$$q = \delta r t a,$$

where δ is a constant (the diffusivity, or co-efficient of inter-diffusion) the magnitude of which depends upon the nature of the gases.

If r, t, and a are each unity, we get

$$q = \delta,$$

and so we define the diffusivity as *the quantity of the substance*

which passes per unit of time through unit area across which the rate of variation of density of the substance per unit length is unity. The quantity r is generally called the 'concentration-gradient,' and may be written in the form $d\rho/dx$, where ρ is the density and x is measured along the line drawn in the direction in which the diffusion is taking place.

We shall find afterwards that the kinetic theory of gases leads to the conclusion that the co-efficient of interdiffusion of gases should be approximately in inverse proportion to the geometrical mean of the densities of the two gases under one atmosphere of pressure. The figures in the first column of the accompanying table give relative values of δ for pairs of gases, the relative values of the reciprocals of the geometrical means of which are given in the second column:

Carbonic Acid and Air	1 ...	1.
Carbonic Acid and Carbonic Oxide ...	1 ...	1.
Carbonic Acid and Hydrogen ...	3·9 ...	3·8.
Carbonic Oxide and Hydrogen ...	4·6 ...	4·8.
Oxygen and Hydrogen	5·2 ...	4·5.

The greatest deviation occurs with oxygen and hydrogen. This is probably due to molecular action between these gases.

110. *Effusion.*—The phenomena presented in the passage of gases through the pores of solids are of great interest, and have been elaborately investigated by Graham. The simplest results are obtained when the solid is practically of infinite thinness and is non-porous, but has a small hole drilled through it. If a gas is kept under constant pressure at one side of the solid, while a vacuum is preserved at the other side, the process of passage of the gas is called *effusion.* The theoretical treatment of the question is extremely simple. The work done in the transference of unit volume of the gas is (§ 62) numerically equal to p, the pressure; for the total pressure on unit area then acts through unit distance. And the kinetic energy acquired is $\frac{1}{2}\rho v^2$, where ρ is the density or mass per unit volume, and v is the speed of the escaping gas. Hence the speed of escape, and therefore the quantity of the substance which passes through in one unit of time, is inversely as the square root of the density. The observed and calculated quantities for four substances are given in the subjoined table:

	Observed.	Calculated.
Carbonic Acid	0·835 ...	0·809.
Oxygen	0·952 ...	0·951.
Air	1 ...	1.
Hydrogen	3·623 ...	3·802.

It appears that hydrogen and carbonic acid pass through more rapidly and more slowly respectively than the above law would indicate. The reason for this will be seen in next section.

111. *Transpiration.*—When the non-porous septum, above referred to, is not thin, the small aperture becomes a tube of exceedingly fine bore, and the gas passes through by *transpiration.* Graham found that the rate of passage was altogether independent of the nature of the substance forming the walls of the tube. This suggests that a layer of the gas becomes deposited upon the interior of the tube, so that the gas has really to flow through a tube composed of its own substance in a highly condensed state. [It is well known indeed that most, and probably all, solids have a great power of condensing gases on their surfaces or within their pores.] Hence we would expect that transpiration is a process which depends upon the viscosity of the gas. This is borne out by the fact that the rates of transpiration of oxygen, air, carbonic acid, and hydrogen are, in increasing magnitude, in the order in which these gases are here named, being fully twice as great in hydrogen as in air—an order which is the exact reverse of their order as regards viscosity. Hence we conclude that the abnormality in the rates of effusion of hydrogen and carbonic acid was due to viscosity, the hole in the thin plate acting to some extent as a short tube.

112. When the pores of a substance through which a gas passes are extremely fine (as in fine unglazed earthenware), the rate of passage follows the ordinary law of diffusion or effusion, *i.e.*, gases pass through at rates which are inversely as the square roots of their densities. Hence we have a means—known as the method of *Atmolysis*—by which to separate a mixture of two gases of different densities. If the mixture be placed inside a porous earthenware vessel, the less dense gas passes through most quickly, so that, when the process has gone on for some time, we have two portions of gas, one containing in most part the less dense gas, the other composed mostly of the denser one. The process may be re-applied so as to separate the two constituents to any desired extent.

It has been already mentioned (§ 79) that carbonic oxide passes rapidly through red-hot iron; and hydrogen passes through palladium, and even platinum, at ordinary temperatures.

In some cases the gas combines chemically with the substance on one side, diffuses through it, and is given off on the other side. This occurs with india-rubber.

CHAPTER X.

PROPERTIES OF LIQUIDS.

113. *Compressibility.*—Liquids, like gases, convey sound, and are therefore compressible and elastic. But they differ from gases, in that their compressibility is usually extremely small. They differ, also, as widely in respect of the law of compression. An inspection of the diagram of § 278 will show that a vapour such as carbonic acid becomes more and more compressible as it approaches the liquefying stage, while, during liquefaction, the compressibility is infinite. The change is in the opposite direction when the whole substance has become liquid; the compressibility is extremely small, and diminishes as the pressure increases. For example, the right-hand portion of the isothermal of 21°·5 is practically a straight line, and therefore the quantity dv/dp is constant. But the compressibility is dv/vdp; and v diminishes as the liquefying stage is approached, so that the compressibility increases. Similar reasoning proves the above statement regarding the liquid condition.

The earlier determinations of the compressibility of liquids were made by means of an apparatus called the *piezometer*, and the more perfect modern appliances all work on the same principle. This apparatus consists of a large glass bulb, having a narrow carefully-graduated stem, which is open at the top. The internal volumes of the bulb and of the stem are accurately measured. The liquid, whose compressibility is to be determined, fills the bulb and part of the stem. A small column of mercury suffices to separate the liquid inside the bulb from water which fills a strong glass vessel, inside of which the bulb is placed. The outer vessel is closed, and pressure is applied by screwing in a plug so as to diminish the internal volume.

The pressure is communicated to the liquid inside the bulb, since there is a complete liquid connection through the stem; and its amount may be measured by means of the compression of air con-

tained in a glass tube, which is closed at the upper end, and is also placed inside the outer vessel.

If the glass were incompressible, the compression of the liquid would be at once found by means of the extent to which the mercury index descended in the stem. If the liquid were incompressible, while the glass was capable of compression, the index would rise. If the liquid and the glass were equally compressible, the position of the mercury would not alter. Hence we see that this experiment really gives the *difference* between the compressions of the glass and the liquid, so that we must first know by experiment the compressibility of the glass of which the bulb is formed. This, as we shall see in the next chapter, is a comparatively simple matter.

Water is compressed by about one twenty-thousandth part of its bulk per atmosphere of pressure added. Unlike all other liquids hitherto observed, its compressibility *diminishes* when its temperature is raised, a minimum being reached about 63° C.

The compressibility of all liquids is lessened by increase of pressure.

114. *Elasticity.*—All liquids, like gases, possess perfect elasticity of bulk, and, in common also with gases, have no elasticity of form.

115. *Viscosity and Viscidity.* — Viscosity is very much more marked in liquids than in gases, and varies greatly from one liquid to another. The slowness of the descent of fine mud in water is due to the viscosity of that liquid, and the slowness of the fall of fine rain-drops is caused by the viscosity of air. Glycerine is one example of an extremely viscous liquid, while sulphuric ether has little viscosity in comparison with it.

One extremely simple method of determining the viscosity of a liquid consists in forcing it under pressure through a cylindrical tube of very fine bore. The quantity which passes through per unit time is directly proportional to the difference of pressure per unit length of the tube and to the fourth power of the radius, and is inversely as the co-efficient of viscosity.

Viscosity diminishes rapidly with increase of temperature.

Viscidity is a related property in virtue of which a liquid can be drawn out into long threads. Other things being equal, a liquid is viscid in proportion to its viscosity; but the molecular forces produce another effect besides viscosity, which acts so as to prevent viscidity (§ 125).

116. *Diffusion.*—Under the diffusion of liquids we include the diffusion of solutions of solids.

Diffusion of liquids is a very much slower process than diffusion of gases. If a solution of bichromate of potassium be carefully introduced at the foot of a vessel containing water, the process of interdiffusion may go on for months before appreciable uniformity is attained.

Many methods (electrical, optical, etc.) exist, by means of which the co-efficient of interdiffusion (the definition of which is identical with that given in § 110) may be determined.

In one method, due to Graham, communication is established between two vessels, each of which contains a liquid capable of diffusing into that contained by the other. Special care is taken to avoid the production of currents whether in the act of establishing communication or because of difference of density of the liquids. The communication is closed after a definite time, and the extent to which diffusion has gone on is determined. A series of precisely similar experiments is made, each experiment of the series lasting for a different interval of time, and the diffusivity is determined from the results.

117. *Osmose.—Dialysis.*—Diffusion of liquids can take place through animal membrane, such as a piece of bladder. The less dense liquid passes through most quickly. If a vessel containing a strong solution of sugar be closed tightly by means of a membranous substance, and then be immersed in a vessel of water, the contents of the inner vessel will rapidly increase, and may finally cause the membrane to break. The process of such transference is called *osmose*.

Liquids may be broadly divided into two classes with reference to the readiness with which they pass through animal membranes. Crystalloid substances, such as common salt, sugar, etc., pass easily through when in solution; but solutions of colloid substances, such as glue, can scarcely pass at all. This is the basis of the process of *dialysis*, which is used for the separation of a mixture of colloid and crystalloid bodies. The mixture is separated from pure water by a portion of animal membrane, through which the substances pass in very disproportionate quantities. One or two repetitions of the process are sufficient to practically separate the two constituents of the mixture.

The method is essentially analogous to the method of atmolysis, which was described in § 112.

118. *Cohesion.*—Cohesion is that property which, apart from the mere gravitation of the parts as a whole, results in the clinging together of portions of matter whether of the same or of unlike kinds. It may be regarded as a result of the so-called molecular

forces. (See, again, § 83.) When a body has been pounded down, so that its parts have been separated beyond the range of the molecular forces, cohesion may be brought about again by the application of pressure sufficient to place the molecules once more within the range of their mutual forces. In the case of a liquid, it is sufficient to merely place the separated parts in contact. (For further treatment, see under Properties of Solids.)

119. *Capillarity.*—It is a well-known law of hydrostatics that the pressure (§ 75) has the same value at all points of a fluid which are at the same level, so that we should expect that the level must be the same at all surfaces of a continuous fluid mass which are exposed to the atmosphere. Yet, in some cases, this is far from being the fact.

If a fine capillary tube be inserted in some liquids, the leve is higher inside the tube than it is outside; while in other liquids

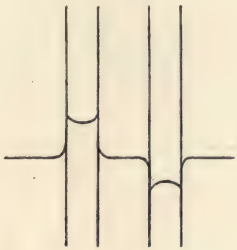


FIG. 63.

the reverse is the case (Fig. 63). Thus water rises inside a glass tube, while mercury descends.

These phenomena (called *capillary* phenomena) seem to be in direct violation of the above-mentioned law of hydrostatics, but in reality they are in strict accordance with it.

120. In proof of this we observe, first, that the surface of a liquid which *rises* in a capillary is always *concave* upwards, while the surface of one which *descends* is invariably *convex* upwards.

Next, we observe that, if a surface, originally plane, is under tension and is curved, there must be more pressure on the concave than on the convex side. Otherwise, the surface would once more become plane, because of its tendency to shrink. Hence, if we can show that there is tension in the surface films of liquids, it follows that there is more pressure on the concave side than on the convex side of the curved surface of a liquid.

But it is well known that the surfaces of liquids tend to become as small as possible. Many examples of this fact may be brought forward. A soap-bubble contracts of itself if the air inside it be in communication with that outside; and the mere fact that the soap-bubble is naturally spherical constitutes another proof, for the sphere is the minimum surface which can enclose a given volume. For the same reason rain-drops are spherical—which is proved by the perfect circularity and definiteness of the rainbow. Again, if some alcohol be dropped on a thin layer of ink, the surface of the ink will decrease, while that of the alcohol is increased, because of the excess of the surface-tension of ink over that of alcohol.

Let us suppose, now, that the surface of a liquid in a narrow tube becomes hollow upwards. Just underneath the outside plane surface there is atmospheric pressure, while just below the curved surface the pressure is less. At some point, p (Fig. 64), lower down in the tube

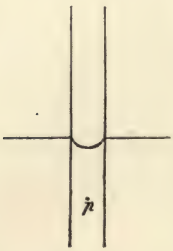


FIG. 64.

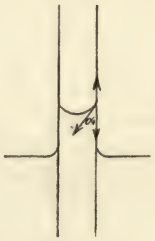


FIG. 65.

the pressure (which is there increased because of the weight of the liquid) is equal to that of the atmosphere. Hence, *in strict accordance with hydrostatic laws*, the liquid *must* rise until the point p is at the level of the surface of the liquid outside.

Similar considerations explain the depression, in a narrow tube, of a surface which is convex upwards.

It only remains to explain *why* the surface becomes curved. We shall assume that the tube is of glass, that the liquid is water, and that the surrounding atmosphere is air. The surface of the liquid in contact with the glass is under tension, the amount of which per unit breadth of the film we may denote by ${}_{w}T_{g}$. There is also a film of condensed gas on the surface of the glass, the tension of which per unit breadth we may similarly represent by ${}_{a}T_{g}$. The particles of the liquid at the edge will therefore be pulled upwards or

downwards according as ${}_aT_g$ is greater than, or less than, ${}_wT_g$. In the case assumed, ${}_aT_g$ is greater than ${}_wT_g$, and so the water surface becomes concave upwards. The tension ${}_aT_w$ of the water surface which is in contact with the air—which formerly acted straight outwards from the walls of the tube—now acts downwards at an angle α with the side of the tube (Fig. 65). So we have now a total downward tension ${}_wT_g + {}_aT_w \cos \alpha$ per unit breadth, and equilibrium will ensue when

$$ {}_wT_g + {}_aT_w \cos \alpha = {}_aT_g. $$

121. In the case of water this equation is not satisfied even when α vanishes, for ${}_aT_g$ is greater than the sum of ${}_wT_g$ and ${}_aT_w$; and so the surface of water in a narrow tube is hemispherical. It is very essential to keep the surface of the liquid free from impurity, and to ensure that the surface of the solid is chemically clean. The slightest trace of grease might entirely prevent the liquid from rising.

The angle α, which is called the *angle of contact*, may be found experimentally by the following method. Let AB (Fig. 66) be a plane plate of the glass (or other solid), and let it be dipped into the

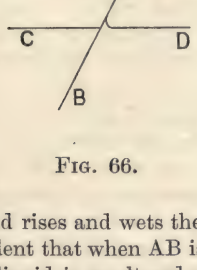


FIG. 66.

liquid CD. If the liquid rises and wets the solid, making an acute angle α with it, it is evident that when AB is inclined at the angle α to CD, the level of the liquid is unaltered at that side of the plate which faces upwards. We may now find α by direct measurement. The figure, if turned upside down, corresponds to the case of a liquid which descends in a capillary tube.

122. We can now determine the height to which a liquid will rise in a tube of given bore. Let r be the radius of the tube, and let T be the tension per unit breadth of the surface separating the liquid from the air, while α is the angle of contact. T cos α is the

upward pull per unit breadth, and hence $2\pi r$ T cos a is the total upward pull. This is balanced by the weight of the raised liquid. Therefore, h being the mean height to which the liquid rises over the outside level, while ρ is the density of the liquid, and g is the value of gravity, we have

$$2\pi r \text{ T} \cos a = \pi r^2 h \rho g.$$

This gives

$$h = \frac{2\text{T} \cos a}{\rho g r}.$$

Hence *the height is inversely proportional to the radius of the tube.*

When the liquid rises between two parallel plates of breadth b, placed at a distance d apart, the above equation becomes

$$2b\text{T} \cos a = bdh\rho g.$$

So in this case

$$h = \frac{2\text{T} \cos a}{\rho g d};$$

that is, while the law is the same as formerly, the height will only be the same when the distance between the plates is equal to the radius of the circular tube.

We can determine the value of T by either of the above methods, provided we know that of a.

The height to which the liquid rises is inversely proportional to d. And since, if the plates be not parallel, but be placed in contact along one of their vertical edges, d will be proportional to the distance from the common edge, the liquid, rising highest where d is smallest, will meet each plate in a curve which is a rectangular hyperbola. The axes of the hyperbolas will coincide respectively with the common edge, and with the lines in which the level surface of the liquid meets the plates.

123. The results of § 120 enable us to explain the strong 'attraction' of two parallel glass plates between which a drop of water is placed. For, since the water becomes concave outwards, the pressure inside it is less than that of the atmosphere, and hence the plates are *pressed*, not attracted, together. The lifting of a stone by means of a leather 'sucker' is similarly explained.

The plates would be apparently repelled apart if the liquid did not wet them, *i.e.*, if it became convex outwards. And it is easy to prove that this would result also if only one plate were wet. The liquid would rise to a greater height outside the one plate, and would descend farther outside the other, than it would rise or fall in the interior space. At the part *ab* (Fig. 67) there is less than

atmospheric pressure outside (for the liquid is concave upwards), while, inside, there is atmospheric pressure. Also, at *cd*, there is more pressure on the inner, than on the outer, side. Thus, from both causes, the plates are pressed apart.

124. In the proof of § 73 we may suppose that, instead of a cord stretched in a circular tube, we have a film of unit breadth stretched over a cylinder. The pressure per unit area will therefore be T/R.

If the film be stretched over a spherical surface of the same radius R, the pressure would have the value 2 T/R, for there are now equal curvatures in two directions at right angles to each other. [The investigation of § 122 furnishes a special proof of this. For if $\cos \alpha = 1$, the surface of the liquid is hemispherical in the circular tube, and is cylindrical in the space between the parallel plates. But the formulæ show that, if the radii of the

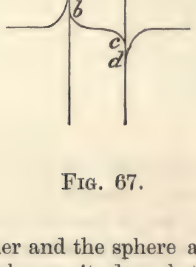


FIG. 67.

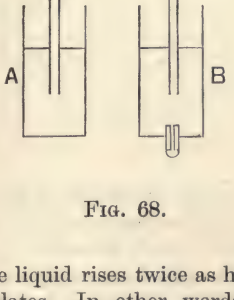


FIG. 68.

cylinder and the sphere are equal, the liquid rises twice as high in the tube as it does between the plates. In other words, the pressure towards the centre of curvature of the film, which supports the weight of the elevated liquid, is twice as great in the first case as in the second.]

In a soap-bubble, we must remember that the liquid film has *two* surfaces; so that, when the bubble is spherical, the air in the interior is subjected to a pressure which is greater than that on the outside by the quantity 4T/R.

The above considerations indicate a method, due to Sir W. Thomson, of measuring the value of T. A capillary tube is inserted in the bottom of a vessel B (Fig. 68) which is partly filled with the given liquid, and is connected by a syphon with a vessel A which also contained the given liquid. By raising or lowering A, the level

of the liquid in B may be altered at pleasure. The liquid will pass through the capillary tube, and will gather into a drop at the lower end of it; but this drop will not fall away unless the difference of level, h, between its lowest point and the free surface of the liquid in B is too great. The inward pressure per unit surface of the drop is $p + 2T/r$, when p is the atmospheric pressure and r is the radius of the drop (measured by micrometric methods). The outward pressure per unit surface, due to the weight of the liquid and the pressure of the atmosphere, is $p + hs$, where s is the specific gravity of the liquid. Hence $2T = rhs$.

125. We have hitherto regarded surface-tension as an observed fact merely, but it is easy to see that it is a necessary result of the mutual potential energy of molecules, or, as we may put it, of the molecular forces. Let p' (Fig. 69) be a molecule in the liquid, situated at a greater distance from its surface than the range through which the molecular forces are sensible. Draw a sphere from p' as

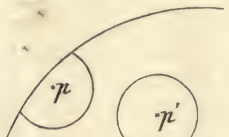


FIG. 69.

centre with the range of the molecular forces as radius. There is no mutual action between p' and any molecule outside this sphere; and it is equally attracted on all sides by the molecules inside the sphere. But any particle p, which is nearer the surface of the liquid than the given distance, is pulled inwards on the whole by the molecular attraction of the interior particles. And this inward pull on the surface particles produces the same effect as, and will obviously be manifested as, a surface-tension, tending to diminish the external periphery of the liquid.

126. The tension of a sheet of india-rubber increases in proportion to the augmentation of the surface, but the tension of a liquid film remains absolutely constant (at least through extremely wide limits) when the area of the surface is altered. If left to itself, the india-rubber will contract until the area of its surface once more attains its original value; but the liquid will contract until its surface becomes as small as possible.

Consider a film of breadth b, the tension of which per unit

breadth is T, so that the total tension is Tb in the direction of the length of the film. If the length of the film be increased by the amount l, the work done in the process is Tbl (§ 62). But this is equal to TS, where S is the increase of surface. Thus the work is directly proportional to the increase of area, and we may look upon T as the amount of work done per unit increase of area instead of a tension per unit breadth.

Taking this fact in conjunction with the result of last section, we can now obtain an expression for the exhaustion of potential energy of molecular separation (the work done by the molecular forces) when two separate masses of the same liquid are placed in contact over a given area. Let S be the area, so that 2S is the diminution of surface of the two masses. The work done is 2TS.

Let T and T′ be the surface-tensions of two different liquids, and let t be the tension of the surface separating the two when placed in contact. The work done by the molecular forces when they come into contact is obviously $(T + T' - t)$ S. The above result is a particular case of this, for, when the two liquids are identical, we have $t = o$ and $T = T'$.

If in any case the work done is greater than $(T + T')S$, *i.e.*, if t be negative, the surface of separation of the two liquids must increase. This it may do by becoming puckered; and, the smaller the scale of the puckering, the greater will be the increase of surface. Thomson regards this invisible replication of the separating surface as the commencement of the process of diffusion.

127. The surface-tension of a liquid diminishes rapidly with rise of temperature, and it vanishes entirely at the critical temperature (§§ 24, 278).

The saturation pressure (§ 275) of the vapour of a liquid depends upon the temperature. But, the temperature being fixed, it also varies with the curvature of the surface of the liquid, being greater the more convex outwards the surface is. Hence small drops of water in a cloud evaporate, the vapour being deposited upon the larger ones.

From the fact that the surface-tension of liquids decreases with rise of temperature, we might deduce, by the principle of stable equilibrium (§ 15), the result that sudden extension of a film will produce a fall in its temperature. For, since the system is in stable equilibrium, it follows that extension of the film will produce an effect which results in an increase of the force resisting the extension. It will, therefore, cause diminution of temperature. This is known to be the case.

The principle of dynamical similarity shows at once that the square of the fundamental period of vibration of a (weightless) liquid sphere is directly proportional to the density of the liquid and to the cube of the radius of the sphere, and is inversely proportional to the surface-tension. It also shows that the period of fundamental vibration of a (weightless) soap-bubble is independent of its linear dimensions.

CHAPTER XI.

PROPERTIES OF SOLIDS.

128. *Compressibility and Rigidity.*—The compressibility of a solid is defined in precisely the same way as that in which we have already defined the compressibility of a liquid or gas. It is the ratio of the percentage change of volume to the change of pressure which produces it. The reciprocal of this quantity is called the resistance to compression, and is usually denoted by the letter k.

The compressibility is most readily determined by measurement of the alteration of length of a rod of the substance to which known hydrostatic pressure is applied. If p' be the percentage alteration of length, the percentage alteration of bulk is approximately $p = 3p'$. For l, b, and t, representing respectively the length, breadth, and thickness of a rod of the given substance, the new volume is $lbt(1-p')^3$, which approximately is $lbt(1-3p') = lbt(1-p)$. In all actual cases the value of p' is so small that any power higher than the first may be neglected. [It must be observed that we assume the substance to be *isotropic*, *i.e.*, its properties are independent of direction. If this were not so, p' might have different values in different directions. This assumption will be adhered to throughout the chapter.]

The rigidity of a solid is the measure of its resistance to change of shape. Let ABCD (Fig. 70) be a cube of a given solid, the edges of

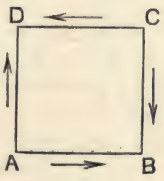


FIG. 70.

which are of unit length. Let equal tangential forces, of magnitude T per unit area, be applied to the opposite faces AB and CD, and

let them act in the directions indicated by the arrows. These forces will produce shearing of the cube, but they will also produce rotation in a direction opposite to that of the hands of a watch. To prevent this rotation, tangential forces equal to the former may be applied to the opposite faces AD and CB. The result of the application of this set of forces is that the square section shown in the figure becomes rhomboidal, the angles at D and B being made less than a right angle by the same amount that the angles at A and C are made larger than a right angle. If θ be the change of angle, the rigidity, which is usually denoted by the letter n, is given by the formula

$$n = \mathrm{T}/\theta.$$

129. Let us denote the three pairs of parallel faces of a unit cube by the letters A, B, and C. Similarly we shall speak of the edges joining the A faces as the A edges, and so on.

Let unit normal pressure per unit area be uniformly applied to the A faces. This will diminish the A edges by an amount l, and will increase the B and C edges by a common amount l'. Now let unit normal tension per unit area be applied to the B faces. This will increase the B edges by the amount l, and diminish the C and A edges by the amount l', small quantities of the second order of magnitude being neglected. The result is that the A edges and the B edges are respectively diminished and increased by the amount $l+l'$, while the C edges are unaltered in length. Hence there is no alteration of volume.

Now this result might have been produced by the method of last section. If any point in the face DC (Fig. 71) of the unit cube be slid forward relatively to a point in AB through the (very small) distance s, the increase of length of the diagonal DB is $s \cos 45° = s/\sqrt{2}$.

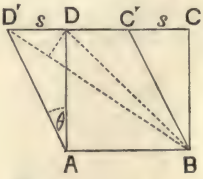


FIG. 71.

Similarly the decrease of length of AC is $s/\sqrt{2}$. The given tangential forces are obviously equivalent to a pressure parallel to AC of magnitude $2\mathrm{T}\cos 45° = \mathrm{T}\sqrt{2}$ per area $\sqrt{2}$, *i.e.*, T per unit area, and to a tension parallel to DB of the same magnitude. If,

now, we let T = 1, $s/\sqrt{2}$ is the alteration of length of the diagonal which contains $\sqrt{2}$ units, and so $s/2$ is the percentage change of length in the direction of the diagonals due to unit tension parallel to BD and unit pressure parallel to AC. Hence, equating the results of the two methods, we get $s = 2(l+l')$. But $s = \theta$, the change of angle of the unit cube. Hence

$$l+l' = \frac{1}{2n} \ldots\ldots (1).$$

130. Unit pressure per unit area on the A faces shortens the A edges by the amount l, and increases the B and C edges by the common amount l'. The quantities l and l' being extremely small, if unit pressure be now applied to the B and C faces, all the edges of the cube will be diminished by the amount $l-2l' = p'$. Hence

$$l-2l' = p/3 = 1/3k \ldots\ldots (2).$$

The quantity $1/l$ (the reciprocal of the percentage change of length of a rod under unit tension or pressure per unit of its transverse sectional area) is called *Young's Modulus*. From (1) and (2) we find

$$l = \frac{1}{9k} + \frac{1}{3n} = \frac{3k+n}{9kn} \ldots\ldots (3).$$

This formula enables us to determine the value of either l, k, or n, provided that we know the values of the other two. The following table gives the value of l for a few substances, the unit of pressure per square inch being the weight of one pound:

Steel	$30(10)^{-9}$.
Iron	$39(10)^{-9}$.
Copper (hard)	$56(10)^{-9}$.
,, (annealed) ...	$64(10)^{-9}$.
Glass (average) ...	$141(10)^{-9}$.

131. The rigidity n is not found in practice directly by the process which was described in § 128. It may be found by determining the moment of the couple which is required to twist a cylindrical rod of the substance through a given angle.

Consider a circular ring of this rod, of radius r, of infinitesimally small breadth dr, and of thickness $dh = dr$. If the ring be divided into little parts by a series of planes passing through the axis OP of the rod, and making angles equal to dr/r with each other, each little part will be cubical in shape, and the number of cubes will be $2\pi r/dr$. If, by the given twist of the rod, the upper side of any one

of these little cubes is twisted forward relatively to the under one by the small angle $d\theta$, the consequent change of angle of the little cube is $rd\theta/dh$. And, as the same change of angle would be produced in

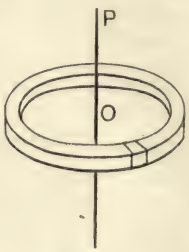


FIG. 72.

a unit cube by a tangential stress of the same magnitude per unit area as that which acts upon the elementary cube, we have (§ 128)

$$\frac{rd\theta}{dh} = \frac{T}{n}.$$

Hence the total tangential stress acting on the circular ring, which is the product of T into the area of the flat surface of the ring, has the value

$$n\frac{rd\theta}{dh}2\pi rdr.$$

The moment of the force on the ring is therefore

$$2\pi n\theta r^3 dr,$$

where θ is the twist per unit length of the rod, so that $d\theta = \theta dh$. Taking the integral of this from the axis outwards, we find for the total moment of the force required to twist the rod through an angle θ per unit of length the quantity

$$\tfrac{1}{2}\pi n\theta r^4, \quad \ldots\ldots (1)$$

r being the radius of the rod.

The quantity $\pi nr^4/2l$, where l is the length of the rod, is called the 'torsional rigidity' of the *rod*. It is not, of course, a specific property.

[It is easy to deduce the result (1) by elementary considerations. Let us imagine the rod to be divided into similar little cubes as above. The total tangential force (in a plane perpendicular to the

axis of the rod) on one face of any such cube, at a distance r from the axis, is proportional to r^2; and therefore the moment of the force is proportional to r^3. But the amount by which, with a given twist, the upper face of the cube slides forward relatively to the lower face is proportional to r. Hence the total couple, c, required to produce the twist is proportional to r^4. But it is also proportional to θ so long as Hooke's Law holds; and we may therefore write

$$c = \tfrac{1}{2}\pi n \theta r^4$$

where n is the rigidity, provided that we make a suitable definition of the various units involved.]

Even a simpler experimental method consists in attaching to one end of the rod a body whose moment of inertia about the axis of the rod is very great. Let the rod be firmly clamped in a vertical position by its upper end, the body being attached to its lower end. The time of oscillation of the whole system about this axis depends upon the value of n.

The moment of the couple about the axis is $I\ddot{\theta}$, I being the moment of inertia of the system (§ 70). Hence, from (1),

$$\ddot{\theta} = \frac{\pi n r^4}{2I}\theta \quad \ldots\ldots (2).$$

This equation asserts that the angular acceleration is proportional to the angular displacement, and therefore the integral is (§ 51)

$$\theta = \theta_0 \cos\left(\sqrt{\frac{\pi n r^4}{2I}}\, t + \alpha\right) \quad \ldots\ldots (3),$$

α and θ_0 being constants. If we increase t by the constant quantity

$$2\pi\sqrt{\frac{2I}{\pi n r^4}},$$

the value of θ is unaltered. This means that the periodic time of vibration of the system is

$$T = 2\pi\sqrt{\frac{2I}{\pi n r^4}},$$

which furnishes a ready method of determining n. (It is here assumed that the rod is of unit length. If it be not so, the equation will still hold, provided that we divide the observed value of T by the square root of the length.)

132. The following table gives, in the same units, the values of n for the substances for which l was given in § 130. From these numbers the values of k are calculated by means of (3) § 130, and

the observed values of k are given in the last column. There is considerable discrepancy between the calculated and the observed values, but this need not produce surprise, for there are many causes of variation in the experimental results. The value of r in the expression for the rigidity is usually small, so that a large percentage error may occur in its measurement; and, even if r were uniform throughout the rod, which is rarely the case—four times this error will be produced in the calculated values of n, since r is involved to the fourth power. Again, the substances may not be really isotropic; and the special physical treatment—*e.g.*, the drawing out of a wire—to which an originally isotropic substance is subjected will frequently make it non-isotropic.

	n	k ($calc^d$)	k (obs^d)
Steel ...	$121(10)^5$ ...	$450(10)^5$ ...	$284(10)^5$.
Iron (wrought) ...	$112(10)^5$ ...	$120(10)^5$ ...	$213(10)^5$.
Copper	$64(10)^5$ to $71(10)^5$ ...	$122(10)^5$ to $288(10)^5$ ...	$227(10)^5$.
Glass (average) ...	$28{\cdot}4(10)^5$ ...	$47(10)^5$ ...	$43(10)^5$.

Amagat's observed values of k for steel, copper, and glass are respectively $220(10)^5$, $174(10)^5$, and $66(10)^5$.

133. The *flexural rigidity* of a bar in a given plane is measured by the moment of the couple which is required to produce unit curvature in that plane. In similar and equal bars of different substances it is directly proportional to Young's modulus; and, in different bars of the same substance and of similar though unequal section, it is proportional to the square of the sectional area. However a bar be bent, the locus of the centres of inertia of all the transverse slices is unaltered in length. The length of all other lines is either increased or diminished. This shows why Young's modulus is involved.

Consider a rectangular rod, of thickness d and of breadth b. Let it be bent uniformly to unit curvature, and let us suppose that b and d are small in comparison with the radius of curvature. If we further imagine the rod to be composed of a very large number of rods, whose cross-sections are similar to that of the given rod, and whose lengths are equal to the length of the given rod, it is easy to see from similarity that the couple which is required to produce the curvature must be proportional to b, the breadth of the rod measured perpendicular to the plane of bending. Also the elementary rods which are further away from the centre of curvature than the central plane of the given rod are extended in proportion to their distance from that plane, while those nearer to the centre of curvature are shortened in the same proportion. Hence we see that the total force must be proportional (so far as

this effect goes) to d and to m—Young's Modulus; and therefore the moment of the couple must be proportional to m and to d^2. But the number of little rods, of a given size, is proportional to d; and therefore, finally, the moment c must be proportional to b, to m, and to d^3—say

$$c = fmbd^3,$$

where f is a constant. Thus we see that the flexural rigidity of a rectangular bar is proportional to its breadth and to the cube of its thickness in the plane of bending.

134. *Elasticity.*—All solids possess elasticity, both of form and of bulk, to a greater or less extent. Within limits (which vary greatly in different substances) the elasticity is perfect, *i.e.*, the original form or volume is entirely regained; but if too great stress be applied, the body will either break or become temporarily or permanently distorted. Steel is a good example of the former class. Its limits of elasticity are very wide apart, and it breaks before much permanent distortion is apparent. On the other hand, lead can scarcely recover entirely from any distortion however slight. When permanent distortion occurs, the molecules have set themselves into new permanent groupings. When the distortion is only temporary, they can resume gradually their old positions.

An elastic solid, if it be kept distorted for a considerable time and then be released so slowly that it does not vibrate, does not in general at once recover its original form, but gradually creeps back to it. If it be consecutively distorted, first in one sense and then in the opposite, it will slowly recover from the second distortion for a time, and then will undo the quasi-permanent part of the first.

The limits of elasticity depend to a large extent upon the physical treatment to which the substance is subjected. The elasticity of a wire is greatly diminished if the wire be kept oscillating for a long period of time. That this is so is shown by the fact that its oscillations die away much more rapidly in this case when it is left to itself after being set in oscillation. The elasticity is said to be 'fatigued' by the process.

135. We may conveniently define elasticity as that property in virtue of which stress is required to maintain strain (§ 68). Within the limits of elasticity, the necessary stress is proportional to the strain. This is known as *Hooke's Law*, and is usually stated in the form '*Distortion is proportional to the distorting force.*'

The constancy of the quantities n and k in equations (1), § 129, and (2), § 130, depends upon the truth of this law; for their constancy implies that, if the unit of pressure be multiplied in any

proportion, the quantities on the left-hand side of the equations will also be increased in the same ratio.

The constancy of the pitch of a note which is given out by a musical instrument proves that Hooke's Law is obeyed by the vibrating substance within the given limits of distortion.

The rigidity and the resistance to compression will not remain unaltered if the distortion be too great; but, between new limits, the law will again hold, n and k having new constant values.

136. *Viscosity.*—Viscosity, or internal friction, is apparent in solids as well as in fluids. The vibrations of a tuning-fork die away at a geater rate than can be accounted for by the impartation of energy to the air in the production of sound. Internal friction transforms the original energy partly into the form of heat in the material of the instrument. The phenomenon of fatigue of elasticity is also due to this cause, the internal friction being increased by the process which induces 'fatigue.'

137. *Cohesion.—Tenacity.*—Cohesion is in general much more powerful in solids than in liquids.

The parts of any body are kept together both by the force of cohesion and by mutual gravitation. In small bodies, such as stones, the part played by gravitation is totally negligable in comparison with that due to cohesion. But, in large masses, such as the earth, gravitation has much the larger effect.

The force of cohesion has generally been regarded as a molecular attribute distinct from gravitation. But Sir W. Thomson has pointed out that cohesion may be explained by means of the gravitational law.

If a bar of lead be cut into two parts, such that the freshly-cut surfaces accurately fit each other, the parts will readily reunite by cohesion when the surfaces are brought sufficiently close to each other. Such a phenomenon as this could not occur if matter were continuous and of uniform density throughout. For the range of the molecular forces (§ 145) is so small that only an extremely small amount of matter at one surface of the bar could sensibly attract, according to the gravitation law, a given particle at the opposed surface. But, in order to get comparatively a very large mass at one surface within 'molecular range' of the given particle at the other, we have only to suppose that, as regards density, matter is intensely heterogeneous on an invisibly small scale; and thus the molecular gravitational force might be sufficiently great to account for cohesion.

138. The property of *tenacity*, in virtue of which there is resistance to the drawing asunder of the parts of a body, is obviously

cohesion regarded from a different point of view. We may measure the tenacity of a substance by the tension which is required to rupture a rod or wire of that substance whose cross-sectional area is unity. The following table gives its values, for *sudden* rupture, in a few substances. Rupture will be *slowly* produced by somewhat smaller (frequently considerably smaller) tensions. These numbers, however, can only be used for purposes of rough calculation, as the property varies with the physical treatment and chemical purity of the substance. They represent the number of kilogrammes whose weight will produce sudden rupture of a rod which has a sectional area of one square millimetre.

Lead	2·4 ...	2.
Tin	2·9 ...	3·6.
Gold	28 ...	11.
Silver	30 ...	16.
Copper	41 ...	32.
Iron	65 ...	50.
Steel	99 ...	54.
Oak	7	
Ash	12	

The numbers in the first and second columns refer to the drawn and annealed states of the metals respectively. The tenacity of wood is of course very much smaller when the length of the rod is taken across the grain than when it is taken in the direction of the grain.

CHAPTER XII.

THE CONSTITUTION OF MATTER.

139. Early in the history of science, discussion arose regarding the possibility of the infinite divisibility of matter. A drop of water may be subdivided into smaller drops of water; but this process cannot be indefinitely continued, for a point is ultimately arrived at beyond which subdivision cannot be carried without alteration of the chemical nature of the substance. But the problem with which we are now concerned goes deeper than this: we wish to know whether or not we would ultimately reach an indivisible part, or atom, of matter.

No answer can be given yet to this question, though various hypotheses have been framed regarding the ultimate constitution, or structure, of matter.

One of the most famous of these hypotheses is known as the *Lucretian Hypothesis* of hard atoms.

According to Lucretius, hard atoms exist; and they are indivisible because they are infinitely hard. The reason which he gave for their existence was that there must be a limit to the decay of matter—which he asserted to be a more rapid process than the agglomeration, or building up, of matter. If there were no such limit, all matter would, he said, have disappeared in the course of infinite past ages. His hypothesis may be true, but his assumption, upon which it was based, is not true; for we know that the building up of matter into larger parts is a more rapid process than its disintegration.

Lucretius further asserted that there must be void spaces between the atoms, otherwise motion of solids in fluids, such as that of a fish in water, could not occur. Here, again, his reason is not conclusive, although his conclusion is correct. Motion of solids in fluids could occur even if the hard atoms were closely packed together. This we know by the principle of fluid circulation in

re-entrant paths. But the fact that all matter is compressible shows that there must be intervals between the hard atoms.

140. The atomic hypothesis of *Boscovich* is a mere mathematical device, which enables us to avoid the physical difficulties of the problem. Boscovich looked upon the atoms as mathematical points endowed with inertia and with attractive or repulsive forces varying with the distance. These forces are always attractive when the distance exceeds a certain superior limit, and are always repulsive when the distance falls short of a certain lower limit. They become infinite when the distance vanishes, so that no two atoms can occupy the same position at the same time. This theory may be worked out so as to determine the properties of a continuous medium so constituted. It has recently been developed by Sir W. Thomson.

141. The most recent atomic hypothesis—and the one possessed of the greatest interest at the present day—is the vortex-atom hypothesis of Sir W. Thomson. According to Thomson, matter consists of the rotating parts of a perfect (§ 74), inert fluid, which fills all space.

Since the fluid is perfect, any part of it, being once set in motion, will remain in motion for ever, and will be completely distinguished from all other parts. Thus the principle of conservation of matter is an essential part of the hypothesis.

The ultimate parts or vortices are indivisible, not from being perfectly hard but, because it is impossible to get at them in order to divide them.

The properties of such vortices may be illustrated by means of smoke-rings, such as those which are occasionally seen issuing from the funnel of a locomotive when the door of the furnace is suddenly closed, or those which are at times blown from the mouth of a cannon. These smoke-rings may readily be produced by means of a box one end of which is flexible and the other end of which is perforated by a circular opening three or four inches in diameter. If a strong solution of ammonia be sprinkled on the floor of the box, and if a vessel containing a strong solution of hydrochloric acid gas (or, preferably, common salt on which strong sulphuric acid is poured) be also introduced into it, dense white fumes of ammonium-chloride are formed. A slight blow on the flexible end of the box is sufficient to drive out a portion of the air from the interior. This part revolves round and round, in the manner which is indicated in section in Fig. 73, and forms a complete circular vortex, the course of which may be traced for some distance through the surrounding still air. As the ring advances, its speed diminishes, and its

diameter increases; and at last it disappears, its motion being stopped by friction.

If a second ring be projected after the former, with slightly greater speed and nearly in the same line, actual collision will not ensue, but each will move aside vibrating as an elastic ring would after direct impact. If it be projected exactly in the same straight line as the first, the second one will grow smaller as they approach, and, its speed increasing, will pass through it. If the relative speed of approach be not too great, the two rings will not separate, but

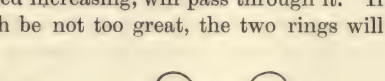


FIG. 73.

will continue to revolve round each other in the manner indicated. This corresponds to molecular or chemical combination of vortex atoms.

Currents are produced in the surrounding fluid which flow forwards in the direction of motion through the interior of the ring and pass back round the outside of it. It is the action of these currents which prevents actual contact between two impinging rings, and which makes it impossible to divide a ring formed in a perfect fluid.

The circular atom is the simplest which can exist, but a vortex-atom may be built up of any number of simple, or knotted, rings, linked, or locked, together in any manner whatsoever.

The mathematical difficulties which beset the investigation of the motion, and the mutual action, of vortices are so great that they have been overcome as yet only to a very slight extent. But, in so far as they have been overcome, the results are not unfavourable to the hypothesis.

The great recommendation of the hypothesis is that it postulates nothing but the inert vortically moving fluid. All the known properties of matter are to be deduced from that one postulate. Other hypotheses assume the existence of special inter-atomic forces, and, if one assumed set fails to produce required results, a new set can be conjured up to suit the case.

Further adaptations of the vortex theory will be given in next chapter, and in Chap. XXXIII.

142. Standing in distinct contrast with these atomic hypotheses, we have the hypothesis that matter is continuous but intensely heterogeneous. The atomic hypotheses may be very roughly illus-

trated by a brick wall built without mortar. The separate bricks represent the various atoms, and there are gaps between them. The hypothesis which we now consider is similarly illustrated by a wall in which the gaps are filled up by cement. Viewed from a distance, the whole seems homogeneous (as matter does even when we inspect it with the most powerful microscopes) but, if sufficiently closely inspected, it is seen to be heterogeneous.

Heterogeneity is a necessity whether atoms exist or not. This is indicated by many physical phenomena, notably by the dispersion of light, which, if the undulatory theory be true, could not occur were matter homogeneous and continuous.

143. If we assume the existence of molecules which exert action in all directions throughout a small range, we can explain the various phenomena of crystalline structure. Let us suppose first that the molecules are in stable equilibrium under their mutual action when their centres are at a definite distance apart. We can obviously make a model of such an arrangement by means of equal spherical balls or marbles, a molecule being supposed to be situated at the centre of each.

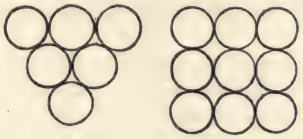


FIG. 74.

We may lay down a plane triangular arrangement of marbles (Fig. 74), and then build up a solid by placing a second set of marbles above the first, so that each marble in the second set fits into the hollow between three in the first set, and so on. In this way we form a regular tetrahedron.

This tetrahedron has six edges, and, if these edges be pared off by planes equally inclined to the adjacent faces, we ultimately get a cube.

The cube has eight vertices, and, if these vertices be pared off by planes equally inclined to the faces which meet at these vertices, a regular octahedron is obtained.

If the twelve edges of the cube be bevelled by planes equally inclined to the adjacent faces, a rhombic dodecahedron is ultimately produced.

It is by no means difficult to account in this way for all the

various symmetrical forms of crystals belonging to the cubic system; and, if we replace the component spheres by ellipsoids of revolution, we can account for all known crystalline forms.

We are thus led to believe that crystalline bodies are built up of particles which set themselves, under the action of their mutual forces, in the position of least potential energy, *i.e.*, in the position of stable equilibrium. The strongest proof of this which we can have is contained in the fact that the lengths of the edges of any given natural crystal are expressible as multiples of *small whole numbers*.

144. The square order of arrangement (indicated in the above figure) in which any one sphere touches four spheres in the layer immediately below, is in no way different from the triangular arrangement just considered; for, if we remove an edge row of spheres from the triangular pyramid, it is at once evident that the particles are arranged in square order in planes which bevel an edge symmetrically.

In both the triangular and the square order of arrangement, any one particle touches twelve others. In the triangular order, a particle touches six others in its own layer, and three in each of the adjacent layers. In the square order a particle touches four in each of these layers.

It is true indeed that all the crystalline forms of the cubic system may be explained by two other methods of arrangement, but these are of no interest to us physically, as they do not correspond to that position of most stable equilibrium which free, mutually attractive, particles would naturally assume. Thus we might build up a cube first of all on the square order of arrangement—a

FIG. 75.

particle in any one layer resting on the top of one in the subjacent layer; and, from this cubic form, we might then, as above, produce the other crystalline forms. Again, we might start with an open square arrangement (indicated in Fig. 75) the particles in any one layer being so far apart that another particle which is set in the hollow between four will just touch a particle which is set in the corresponding hollow on the other side of these four. In this way,

also, all the crystalline forms of the regular cubic system may be produced. In the former method, any particle touches four particles in its own layer, and one particle in each of the adjacent layers—six in all; in the latter method, a particle touches none in its own layer, four in each of the layers immediately preceding and succeeding its own, and one in each of the layers outside of these—ten altogether. In neither of these cases is the equilibrium so stable as in the case which was considered in last section.

145. Many other physical phenomena, besides those which are exhibited by crystalline bodies, make it a practical certainty that matter is molecular in its structure; and these phenomena also enable us to obtain approximate estimates of the range of the molecular forces, and of the average distance between contiguous molecules.

One such class of phenomena is that exhibited by liquid films. We have seen (§ 125) that the existence of molecular forces would account for surface-tension. And we have seen, also (§ 126), that the surface-tension remains practically constant until the thickness of the film is very largely reduced. There are optical methods (§§ 218-221) by which the thickness can be very accurately ascertained. Reinold and Rücker have shown by these methods that the surface-tension of a soap-bubble begins to diminish when the thickness is between 96 and 45 micro-millimetres (millionths of a millimetre, one inch being about equal to 25·4 millimetres). It diminishes to a minimum when the thickness is 12 micro-millimetres, and then increases again to a maximum.

Plateau had previously shown that the tension is unaltered when the thickness is reduced to 118 micro-millimetres, and he concluded that the range of molecular forces is less than 59 micro-millimetres. (Maxwell, however, has given theoretical reasons for the belief that the tension may not alter until the total thickness of the film is equal to the range of the forces, which would make Plateau's superior limit 118 micro-millimetres.)

By measurements of the height to which liquids rise, because of the so-called capillary forces, between parallel glass plates which were coated with very thin wedge-shaped metallic films, Quincke was led to the conclusion that the forces between the glass and the liquid became evident when the thickness of the metallic film was 50 micro-millimetres.

Wiener has found that the phase of the vibrations of light (§ 243) which is reflected from a thin film of silver deposited on mica begins to alter when the thickness is 12 micro-millimetres, and that it was possible to detect the presence of a film of silver not exceeding 0·2 micro-millimetre in thickness.

Estimates of the magnitude of the molecular range have also been based upon the thickness of the films of gas which are condensed upon solids, but these are open to great objection.

From all these results we may conclude that the order of magnitude of the range of molecular forces is about 50 micro-millimetres, or one five-hundred-thousandth part of an inch.

146. It is possible, also, to obtain an estimate of the coarse-grainedness of matter; that is, of the average distance between molecules.

If (see § 324) two plates of different metals, say copper and zinc, be placed in contact, each becomes electrified oppositely, and so they attract each other. If the plates have an area of one square centimetre and be at a distance of one-hundred-thousandth of a centimetre apart, they will, when joined by a metallic connection, attract each other with a force of two grammes weight. Hence the work done in bringing them into this position by means of the electric attraction alone would be 2/100,000ths of a centimetre-gramme. If we now build up a cube of such pieces of metal, alternately zinc and copper, the thickness of each being 1/100,000th of a centimetre, and the distance apart of each pair being 1/100,000th of a centimetre, the work done by the electric attraction is 2 centimetre-grammes. If this work were spent in heating the mass of metal, the temperature would rise by less than the sixteen-thousandth part of a degree centigrade. But if the thickness of the plates and their distance apart were 1/100,000,000th of a centimetre, the work would be sufficient to raise the temperature of the mass by nearly 62° C. And, if the plates and the spaces between them were made yet four times thinner, the work would produce more heat than is given out by the molecular combination of zinc and copper. Hence the magnitude of the contact-electrification of zinc and copper must sensibly diminish before the substances are so finely divided as we have supposed. But this suggests that there cannot be many molecules in a thickness of the 1/100,000,000th of a centimetre—possibly the coarse-grainedness may even be on a larger scale.

We have already seen (§ 126) that the work required to increase the area of a water film by one square centimetre is numerically equal to the tension of the film per linear centimetre. But the magnitude of the tension is about 16 centigrammes weight per linear centimetre. Hence, if we increase the surface of a water film by n square centimetres, we have to do n times 16 centimetre-centigrammes of work. But a water film cools when it is stretched; and Sir W. Thomson has shown that, if the temperature is to be kept constant—which is necessary if the tension is to remain

constant—we must supply heat to the film, to an extent which would require the expenditure of about half as much work again. So, finally, if the area of a water film be increased, at constant temperature, by n square centimetres, work must be expended to the extent of $24n$ centimetre-centigrammes. If we start with a cubic centimetre of water, and increase its area (and therefore diminish its breadth) one-hundred-million-and-one-fold, we must expend 2,400,000,000 centimetre-centigrammes of work. But this work, if spent in heating the liquid, would be more than sufficient to completely volatilize it. Hence, if the thickness can be diminished to this extent, the surface-tension must greatly decrease in magnitude. And so we conclude that there cannot be many molecules of water in a thickness of 1/100,000,000th of a centimetre.

The fact of the dispersion of light in its passage through dense transparent media proves, as has been already stated, the heterogeneity of these media. Starting from certain assumptions regarding the constitution of such media, Cauchy deduced results which indicate that there are only about ten molecules to the wave-length of violet light (about $4{\cdot}10^{-5}$ cm.) in ordinary glass. This cannot be admitted for many reasons. But Thomson has recently shown that it is possible to modify Cauchy's theory in such a way as to widen the limit which it sets.

The kinetic theory of gases gives (§ 153) another, and even more certain, indication of molecular magnitude.

From these four courses of reasoning, Sir W. Thomson concludes that there are not more than 10^9, nor less than $5(10)^6$, molecules per linear centimetre in ordinary liquids or solids.

Another method consists in measuring the thickness of the dielectric layer separating an electrolyte from the electrodes. This thickness is presumably the distance between the molecules of the liquid and the contiguous molecules of the solid electrode. The method gives values of the number of molecules per linear centimetre which vary from 10^7 to $5(10)^8$.

Again, if we suppose a cubic centimetre of any liquid to be divided up by three sets of n planes parallel to the three pairs of faces of the cube, the surface of the liquid is increased by $6n$ square centimetres. The work which is required to produce this increase of surface is (neglecting the equivalent of the heat required to keep the temperature constant) $6n$T centimetre-grammes, where T is the surface-tension expressed in grammes weight per linear centimetre. [It is assumed, of course, that T is practically constant during the process. We have seen that Plateau found it to be constant in a water film down to a thickness of 118 micro-millimetres. And

Reinold and Rücker have shown that the value of the second maximum, which is reached when the thickness is 12 micro-millimetres, only differs from the former value by about 0·5 per cent.] If n be the number of molecules per linear centimetre, the quantity $6n$T is approximately equal to the work required to break up the cube of the liquid into its constituent molecules. It is therefore equal to the work-equivalent of the latent heat (§§ 276, 289) of the liquid. If L be this quantity, we have, therefore, as an approximation, $n = \mathrm{L}/6\mathrm{T}$. The liquids, water, alcohol, ether, chloroform, carbon bisulphide, turpentine, petroleum, and wood spirit, have, according to this method, 50, 52, 30, 15, 19, 30, 40, and 70 millions, respectively, of particles per linear centimetre. These numbers all lie well within the extreme limits given by Thomson. Of course, no stress is to be laid upon the relative, or even the absolute, values of the figures; the point of interest is the close agreement as to the *order* of the unknown quantity.

CHAPTER XIII.

THE KINETIC THEORY OF MATTER.

147. The first glimmerings of the idea that the observed properties of matter may be due to motion occurred as far back as the time of Democritus and Lucretius. But the idea did not develop into an actual physical hypothesis until Hooke, and, later, Daniell Bernoulli suggested that gaseous pressure may be due to the impact of the molecules of the gas upon the sides of the vessel which contains it. Somewhat later, Le Sage, as we have already seen, applied the same principle to the explanation of gravitation, and various developments were made by Prevost and Herapath. In 1848, Joule calculated the speed which the particles of a given gas must have in order to produce a given pressure. But the full mathematical development of the Kinetic Theory of Gases is due mainly to Clausius and Maxwell.

148. In the kinetic theory it is supposed that the particles of a gas are darting about in all directions with great average speed. Some of the particles may, for a short time, have very much smaller speed than this average—may indeed be at rest for a moment; and others may be moving with much greater speed than the average. This average is the square root of the mean of the squares of the individual speeds of the various molecules, and is called the *mean square speed.*

Collisions are supposed to occur amongst these particles. In the interval between any two successive collisions of a particle there is a certain average distance which the particle describes, and this distance is called the *mean free path.* The mean free path, under ordinary conditions, is large in comparison with the diameters of the molecules, which are regarded as being smooth hard spheres with unit co-efficient of restitution.

The collisional force between two molecules is assumed to be repulsive, just as it would be in the case of elastic solids. But, actually, in many, or even most, of the so-called 'collisions,' true contact may not occur. The real forces may be attractive, and

two rapidly-moving molecules coming within range of mutual attraction may whirl round each other in sharply-curved paths, as a comet dashes round the sun, the result being the same as if actual contact had taken place. Some contacts must occur unless the molecules are infinitely small, which we cannot admit.

Experiments made by Joule and Thomson on certain gases show that, if the density of each be varied while its total energy remains constant, the temperature is somewhat higher when the density is greater. Hence, if (§ 150) equality of temperature in two portions of gas means equality of average kinetic energy per molecule, it follows that the potential energy is somewhat less in the denser condition. But this indicates molecular attraction at the average distance of the molecules experimented with.

149. *Gaseous Pressure.—Boyle's Law.*—Let n be the number of molecules per unit volume which are moving in a given direction with speed which differs extremely little from a certain quantity u. The number of such molecules which pass per unit time across unit area drawn perpendicular to the direction of motion is nu; and, if m be the mass of each molecule, the momentum which these particles carry with them is mnu^2. By Newton's Second Law of Motion, this must be equal to the pressure produced by such molecules on the side of the vessel. Hence, *the square of the speed being involved*, we see that, so far as the pressure is concerned, we may assume that each particle is moving with the mean square speed.

The total pressure per unit of area in the direction considered, which we may assume to be that of the x-axis, is therefore $m\mathrm{N}\overline{u^2}$, where N is the total number of particles per unit volume, and $\overline{u^2}$ is the mean value of u^2 for all the molecules. Similarly the pressures per unit area in the direction of the y and z axes may be written $m\mathrm{N}\overline{v^2}$ and $m\mathrm{N}\overline{w^2}$ respectively. But, in a gas which is at rest as a whole, all these quantities are equal; and so we have for the pressure p the expression $p = \frac{1}{3}m\mathrm{N}(\overline{u^2} + \overline{v^2} + \overline{w^2})$, which we may put in the form

$$p = \tfrac{1}{3}m\mathrm{N}\overline{\mathrm{V}}^2 = \tfrac{1}{3}\rho\overline{\mathrm{V}}_2 \ldots\ldots (1)$$

where $\overline{\mathrm{V}}$ is the mean square speed independent of direction and is the density of the gas. By means of this result, Joule calculated the value of $\overline{\mathrm{V}}$ in various gases. In hydrogen it is somewhat over 6,000 feet per second.

We have seen that $\overline{\mathrm{V}}$ is constant when the temperature is steady, so that this equation asserts that the density of a gas is directly proportional to the pressure—which is Boyle's Law.

150. *Avogadro's and Charles' Laws.*—The equation (1) may be written

$$pv = \tfrac{1}{3}\overline{V}^2 \ldots\ldots (2)$$

where v is the reciprocal of ρ, *i.e.*, it is the volume of unit quantity of the gas. If we compare this with the expression (§ 266) $pv = RT$, we see that the mean square of the speed of molecular motion is proportional to the absolute temperature as measured by a gas thermometer (§§ 266, 267) filled with the particular gas under consideration.

It follows from the principles of the kinetic theory that, in a mixture of two gases in equilibrium, the average kinetic energy per molecule of each gas is the same. And, if we assume the truth of Avogadro's law that there is the same number of molecules in unit volume of each of two gases at given temperature and pressure, (1) shows that in two such gases the average kinetic energy per molecule is identical. Conversely, if we assume—as is done in the kinetic theory—that two gases at the same temperature have the same average kinetic energy per molecule, we can deduce Avogadro's law as a consequence. But it must be carefully observed that the truth of Avogadro's law does not establish the truth of this assumption; it merely proves on this theory that two gases at the same temperature *and pressure* have equal average kinetic energy per molecule. If, however, the gases rigidly obey both this law and Boyle's Law, the truth of the assumption follows: for then, in any one gas, $m\overline{V}^2$ remains constant, no matter how much p may vary.

Equation (2) shows that any gas which obeys these laws will expand equally for equal increments of temperature, *provided that the temperature be measured by the average kinetic energy per molecule*, which is a generalisation of the above assumption. And, further, with this proviso, the equation also shows that any two such gases will expand proportionately as their common temperature rises. These two results constitute Charles' Law (§ 265).

The deviations from Boyle's and Charles' Laws can be explained on the kinetic theory if we take into account molecular action between the particles at greater than collisional distances.

151. *Diffusion.—Thermal Conductivity.—Viscosity.*—The question of gaseous diffusion has been already discussed in Chap. VIII. The kinetic theory asserts that each particle is darting about with great speed, and is only prevented from moving rapidly away from the neighbourhood of its position at any given instant by means of the great number of collisions to which it is subjected by other

particles. These collisions change the direction of motion of the molecule an immense number of times per second, and thus the diffusion of particles throughout a mass of gas is a very slow process. The law of inter-diffusion given by the theory is identical with that deduced from observation.

The diffusing molecules carry their kinetic energy with them, and there is interchange of energy during collision. This transference of energy constitutes (Chap. XXIV.) the process of thermal conduction. And this process is slightly more rapid than the transference of the molecules themselves; for, though a molecule may be turned back by a collision, its energy is handed on by means of the molecule which collides with it.

If two portions of gas are moving relatively to each other, the molecules of each inter-diffuse, and consequently there is interchange of momentum. But, on the average, the momenta of the particles of each gas are in opposite directions, and so the relative motion is gradually stopped. This explains gaseous viscosity. An illustration may be taken from two railway trains running on parallel lines past each other. If luggage were thrown constantly from each train into the other, the interchange of momentum resulting from the impacts might soon reduce the trains to relative rest.

152. *Evaporation, Dissociation, etc.*—As a gas becomes more and more condensed the mean free path of its molecules becomes smaller and smaller, until, in the liquid state, its magnitude is excessively small in comparison with the magnitude of the mean free path of a gaseous particle. Still, in the liquid, the molecular action is precisely of the same character as that of a gas. But, as a result of the comparative closeness of the molecules, the transference of energy by impact (that is, the conduction of heat) is a much more rapid process than the transference of the molecules themselves; and the rate of diffusion of the molecules of a liquid is very small in comparison with the rate of diffusion of gaseous molecules.

In a liquid some of the quickly-moving particles may escape from the attraction of neighbouring molecules and become particles of vapour. At the same time some vapour particles may become entangled in the liquid. When these two processes take place at the same rate, the liquid and its vapour are said to be in equilibrium. When the former process occurs most rapidly, the liquid is said to be evaporating; and, when the latter process preponderates, the vapour is said to condense.

Dissociation occurs when an impact is so violent as to break up a compound molecule into its constituent parts. Probably dissocia-

tion occurs in all fluids to a slight extent, even when their temperature is far below the ordinary temperature of dissociation, but, in this case, recombination occurs as rapidly. As the temperature rises, the impacts become more violent and dissociation goes on at a greater rate, but not at so great a rate that recombination cannot occur as quickly. When, finally, the so-called temperature of dissociation is reached, recombination is unable to balance dissociation, and the substance is resolved into its components. If the temperature is again allowed to fall, recombination may, or may not, occur, depending on the condition whether or not energy will be degraded (§ 11) in the process. If more energy will be degraded by the occurrence of recombination than by its non-occurrence, recombination will ensue. [See § 280.]

An important result of the kinetic theory is that, in a vessel filled with a mixture of different gases, the final distribution of each gas under the action of gravity is the same as if no other gas had been present.

Another very important result (which, together with the former, is due to Clerk-Maxwell) is that a vertical column of gas, when in equilibrium under the action of gravity, has the same temperature throughout, or, in other words, gravity has no effect upon the conditions of thermal equilibrium.

153. An expression has been deduced from this theory by which we can calculate the length of the mean free path of a molecule in an ordinary gas, such as air, in terms of observed values of the viscosity, and of the material and thermal diffusivities. According to Maxwell the value of this quantity in the case of hydrogen is $3{\cdot}8(10)^{-6}$ of an inch (roughly four millionths). The theory also shows that the number of particles per cubic inch of ordinary gases is about $3(10)^{20}$ (that is, 300 million million millions), and that the diameter of a (supposed hard and spherical) molecule is about $2{\cdot}3(10)^{-8}$ inch.

The following results are given by Maxwell:

	Hydrogen.	*Oxygen.*	*Carb. Oxide.*	*Carb. Acid.*
Mean square speed at 0° C. expressed in feet per second	6190	1550	1656	1320
Mean free path in thousand-millionths of an inch	3860	2240	1930	1510
Number of collisions per millionth of a second	17750	7646	9489	9720
Diameter per inch $\times (10)^{-11}$	2300	3000	3500	3700.

The length of the free path increases as the density of the gas diminishes. Tait and Dewar first found that, in a good vacuum, it

may amount to several inches. The action of Crooke's *radiometer* depends upon the great length of the free path. [This instrument consists of four vanes of mica mounted at the ends of two light rods, which are fastened by their centres to a vertical axis which is free to rotate. Each disc of mica lies in the vertical plane which passes through the rod to which it is attached, and the two rods are placed at right angles to each other. Also each disc is blackened on one side and is bright on the other, the blackened sides being all similarly situated with regard to rotation around the vertical axis. The whole is mounted inside a glass vessel from which the air is largely extracted. Radiant heat (or light) falling upon the vanes is absorbed more freely by the black surfaces than by the bright surfaces, and so the former become hotter than the latter. Hence the particles of air which impinge upon them are driven off again more violently than are the particles which impinge upon the bright sides, and so, by the third law of motion, the necessary reaction results in rotation.

The exhaustion of the air is carried to such an extent that a particle which is repelled from a disc rarely encounters another particle before it strikes the side of the vessel. If the exhaustion is carried too far the effect is diminished, because the number of particles which strike a disc in a given time is lessened.]

Gaseous matter whose molecules have very large free paths is sometimes called 'radiant' matter.

154. The statement that the diameter of one of these hypothetical (hard, smooth, spherical) molecules is 2·3 hundred-millionths of an inch is not intended to imply that this is the size of an actual material molecule. It merely asserts that, on the average, this is about the least distance to which the centres of two molecules can approach during collision.

But, even if they be smooth and spherical, the molecules of matter cannot be hard; for, in this case, the time of describing the mean free path must vary inversely as the average speed of the particles. But, on the contrary, Maxwell's experiments on gaseous viscosity show that the time is independent of the speed. This would be possible with soft elastic particles.

There are other, even more cogent, reasons why the molecules cannot be smooth, hard, and spherical. Thus the great complexity of the spectra (Chap. XVII.) of many gases and vapours shows that the molecule is a very complex structure with a great many degrees of vibrational freedom. But a hard, smooth, spherical molecule has, in effect, only three degrees of freedom—all translational. And again, the theory asserts that each additional degree of freedom

which the molecule possesses requires that the ratio of the specific heat (§ 271) of the gas at constant pressure to that at constant volume shall be made larger; and the actually observed values of this ratio are far smaller than that indicated by the theory. Further, it seems certain that, in a gas whose molecules are small, elastic, solid bodies, the energy of translation of the molecules must gradually be converted into energy of vibration of smaller and smaller period, so that finally the particles would come to rest.

155. But these difficulties with which the theory is beset do not lead us to discard it. Its results, briefly indicated in part in the preceding sections, place it upon far too firm a basis; and we seek, rather, to inquire more closely into the truth or probability of the fundamental assumptions of the theory, and into the correctness of our deductions from them. The difficulty regarding the specific heats seems to be due largely to a too rapid theoretical generalisation. And the difficulty regarding the transformation of the translational energy of elastic solids into vibrational energy does not, Sir W. Thomson remarks, seem to apply to the case of fluid vortices.

156. We are not, however, to rest content with a kinetic theory of gases only. We wish, if possible, to recognise all the properties of matter as the results of motion of the parts of a medium in which we postulate nothing but incompressibility and inertia.

We can make, from rigid portions of matter, a complex which possesses elasticity. A gyrostat (essentially a heavy fly-wheel, of great moment of inertia, free to rotate about its axis) mounted on gymbals, but not set in rotation, serves very well as a model of a plastic body. It may be turned about into any position which we please, and has no tendency to re-assume its original position. But, if the fly-wheel be set in rotation about its axis, the system at once acts as if it possessed rigidity and elasticity. If it is sharply struck, so as to suddenly turn its axis to a slight extent from its original direction, it will oscillate rapidly about, and finally come to rest in, its first position.

Again, by means of rigid rods on which revolving fly-wheels are pivoted, we can construct a framework, which acts like an elastic spring, and may represent an elastic molecule. By joining together millions of such arrangements, it is possible to construct a model of an elastic solid through which distortional waves will pass, and which, under proper conditions, will exhibit, with regard to these waves, the same phenomena as those which are shown when light passes through a medium placed in a field of great magnetic force.

The vortex theory shows that all that can be done by such a model might be done by a model in which vortices in a perfect fluid take the place of the revolving fly-wheels. And further, with such vortex molecules we can construct—what cannot be done by the rigid fly-wheel molecules—a model of a gas, whose molecules exert mutual force.

Such considerations lead us to believe that we may ultimately be able to explain apparently statical properties of matter in terms of motion.

CHAPTER XIV.

SOUND.

157. The word *sound* is generally used with reference to the physiological effect which results from excitation of the hearing organ. In physical science the term is applied to the external cause of this subjective impression.

We are accustomed to say that sound travels *through* a given medium, whether solid, liquid, or gas; and by this we imply that the particles of the medium do not move forward from the source of the sound to the place at which it is heard. Intermittent impact of such particles would account for the vibration of the tympanum which is necessary to the production of the mental impression of sound. But it is quite obvious that the particles of a solid cannot move forward in the manner indicated, and it would be unscientific to assume that the method of transference of sound in a gas differs totally from the method of transference in a solid body, especially when we consider that sound, after travelling through a solid, may be communicated to, and travel through, a gaseous medium, such as air. We do not, however, need to rely upon any such semi-metaphysical argument, for the speed of sound in air is so great that the forward motion of the particles—did it occur—would correspond to a hurricane of much greater violence than any ever observed. In short, we *know* that the passage of sound through air is not accompanied by such motion—witness the well-known fact that sound is usually best heard in *still air*.

But the passage of sound through air may communicate vibratory motion to objects immersed in it — *e.g.*, the tympanum — and it therefore involves transference of energy, which implies motion of matter. This motion can be nothing else than vibratory motion of the particles of the medium, the state of motion being handed on from particle to particle in the form of a wave which may cause vibratory motion of any solid object which it reaches, just as a wave sent along a stretched cord may cause motion of any object to which the cord is attached.

The vibrations of the particles of a medium may be transverse to the direction in which a wave travels, or they may take place in that direction. A little consideration will show that, in a sound-wave, it is the latter which must take place. Let us suppose that a sound is started by an explosion at a certain point. We know that the sound travels outwards in all directions from this point as centre; and we know also that the particles at the point are driven outwards by the explosion, so as to produce a state of great condensation in the immediate neighbourhood. But, the medium being elastic, the compressed portions expand, and so cause compression of the adjacent parts, and thus the state of compression travels outwards from the centre. Since a state of rarefaction is produced at the very centre of the explosion, the particles in the compressed part just outside rush back so as to fill the partial vacancy, and so the state of compression is succeeded by a state of rarefaction, which travels outwards at the same rate. Thus we see that sound consists in the propagation of a condensational-rarefactional wave, the particles of the medium vibrating to and fro in the direction in which the wave travels.

[It is easy to construct a model, which will roughly illustrate this process, by means of a row of equidistant balls which are attached by strings of equal lengths to a horizontal straight rod.]

The distance, measured in the direction of propagation, from any particle to the next similarly-moving particle, is called the *wave-length*. It may be measured, for example, between points of maximum condensation, or between points of maximum rarefaction.

The actual motion of any particle is a simple harmonic motion, and the terms used in the discussion of such motion (§ 51)—*amplitude*, *phase*, *period*, *etc.*—are therefore used here with similar meanings.

It is customary to speak of a single disturbance propagated through the air as a *noise*. The term *sound* is employed when a periodic disturbance, or series of disturbances, is propagated—the sound being *musical*, or *non-musical*, according as the disturbances are of regular or of irregular period.

When a musical sound is produced, it is usual to speak of the *note* or *tone* which is given out, but it is better to limit the application of the latter term to a simple sound of one definite period only. We shall see subsequently that no sound usually given out by a musical instrument consists of a simple tone only. The term *note* may conveniently be applied to the composite sound which is actually given out.

All sounds differ from each other in three points only—*intensity*, *pitch*, and *quality*. These will be treated in detail afterwards.

158. We shall now proceed to investigate the propagation of sound through any gaseous medium.

For the sake of simplicity, we shall suppose that a disturbance is propagated in the form of a plane wave; that is to say, we assume that any continuous set of particles, the motion of which is in the same phase, lie in a plane.

Let u be the speed with which the sound-wave travels. The actual speed, v, of any particle is the resultant of this speed u and the speed due to the simple harmonic motion of the particle. Hence an ideal plane, which is perpendicular to the direction in which the sound is going, and which moves in that direction with speed u, possesses the characteristic that the instantaneous speed of all particles which cross it is constant. (The value of this speed depends, of course, upon the initial position of the plane.) A statement equivalent to this is that the plane moves so as to be always in a position of constant density.

If ρ be this density, we may write

$$\rho v = c, \quad \ldots\ldots\ldots\ldots (1),$$

where c is a constant; for this equation expresses the fact that the *total* momentum of the particles which cross unit area of the plane in unit time is invariable. But any two such planes, which remain equidistant, obviously contain between them a constant amount of matter which moves with constant total momentum. Hence c is absolutely constant, and so, by the methods of Chap. IV., we get from (1)

$$\frac{d\rho}{dv} = -\frac{c}{v^2} = -\frac{\rho}{v} \quad \ldots\ldots (2).$$

If the pressure per unit area of the plane under consideration be p, the pressure per unit area of another plane at a distance dx from it is $(p + dp/dx \,.\, dx)$. Hence the difference of pressure per unit area of two such planes is $-dp/dx \,.\, dx$. But, by the second law of motion, this must be equal to the instantaneous rate of increase of the momentum of a column of the gas, of unit section, contained at the given instant between these planes. Hence, the amount of matter in this little volume being ρdx, we get $\rho dv/dt = -dp/dx$. That is, $\rho v dv/dx = -dp/dx$, and so

$$\frac{dv}{dp} = -\frac{1}{\rho v} \quad \ldots\ldots\ldots (3).$$

From (2) and (3) we have at once

$$v^2 = \frac{dp}{\rho d} \quad \ldots\ldots\ldots\ldots (4).$$

In a gas which obeys Boyle's Law and Charles' Law, the relation $p = Rt\rho$ holds—R being a constant and t being the absolute temperature. If the temperature be constant, this gives

$$\frac{dp}{d\rho} = \frac{p}{\rho} \quad \ldots\ldots\ldots\ldots (5),$$

and (4) takes the form

$$v^2 = \frac{p}{\rho} = Rt \quad \ldots\ldots\ldots (6).$$

But the compressions and rarefactions which take place when sound passes through a gas take place so rapidly that the equation $p = Rt\rho$ does not represent the actual conditions. Instead of it we must write (§ 302)

$$p = Rt\rho\gamma,$$

where γ is the ratio of the specific heat of air at constant pressure to its specific heat at constant volume; so that, instead of (5), we get

$$\frac{dp}{d\rho} = \gamma\frac{p}{\rho} \quad \ldots\ldots\ldots\ldots (5');$$

and (6) becomes

$$v^2 = \gamma\frac{p}{\rho} = \gamma Rt \quad \ldots\ldots\ldots (6').$$

159. The general equation (4) gives the value of v *under all conditions* of the substance, and therefore (6′) gives the value of v under ordinary conditions of the gaseous medium, provided only that we put the normal values of p and ρ in the expression on the right-hand side of that equation.

Hence we see that, in cases in which (6′) is sufficiently nearly true, the speed of sound is *independent of the pressure;* and that, in all such cases, it is *proportional to the square root of the absolute temperature.*

Also, in different gases under equal pressure and at the same temperature, the speed is inversely proportional to the square root of the density provided that (6) is true. The following table shows how nearly the theoretical relative speeds correspond to the actual relative speeds as observed in a few of the more common gases:

	Observed.	Calculated.
Air	1·000	1·000.
Carbonic Acid	0·786	0·811.
Oxygen	0·953	0·952.
Hydrogen	3·810	3·770.

It must be remembered that these results are calculated on the supposition that the gases are perfect.

[The following considerations enable us to deduce, in a much less strict manner, the general conclusions arrived at by means of the investigation of § 158.

The speed of propagation of a given state of compression depends upon the readiness with which the compressed gas recovers its original condition; and this, in a perfect gas, is proportional to the resistance to compression, which, § 104, is measured by the pressure. But the readiness of recovery depends also upon the mass which has to be moved before the substance can expand, *i.e.*, it depends upon the density, being greater the smaller the density is. Hence we see that, in a perfect gas, the speed of sound depends only on the temperature; for (the pressure remaining the same) the density diminishes as the temperature increases, while, if the temperature be constant, the pressure and density vary proportionately.]

160. Equation (4) gives, not merely the speed of sound in a gas but, the speed of plane sound-waves in any substance. The resistance to compression, k, is by definition, § 104, the ratio of the increase of pressure to the percentage decrease of volume which it produces. But the percentage decrease of volume is equal to the percentage increase of density. Hence we have $k = \rho \, . \, dp/d\rho$, and so (4) becomes

$$v^2 = \frac{k}{\rho} \quad \ldots\ldots\ldots \quad (7).$$

In the case of a perfect gas, $k = p$, and (7) becomes identical with (6).

The speed of sound in water is nearly four times as great as it is in air. This was found experimentally by means of observations made at the Lake of Geneva, the waters of which are comparatively free from currents. A bell was sounded under water, and the mechanism which rang the bell simultaneously fired a gun placed at the surface of the water at the same spot. A large receiver filled with air was placed below the surface at a known distance from the bell. An observer, who listened at the extremity of a tube connected with the receiver, knew when the sound which travelled through the water reached the receiver. He also observed the instant at which the sound of the report of the gun reached him through the air; and so, knowing by the flash of the gun the instant at which both sounds were produced, he compared the speeds in air and water.

In solids the speed is still greater.

The speed in air may be roughly found by observing the interval of time which elapses between the instant of seeing the flash of a distant gun and the instant at which the report is heard. If the air is not still, two simultaneous observations must be made, in one of which the sound travels with the wind, and in the other of which it goes against the wind. The mean of the two results must be taken.

A more accurate method of determining the speed in gases will be given afterwards (§ 172).

161. The loudness, or *intensity*, of a sound is due to the kinetic energy per unit volume of the medium, and so it depends upon the amplitude of vibration of the particles, being proportional to the square of that quantity. For the same reason it depends also, other things being equal, upon the density of the medium, being greater as the density is greater, and vanishing when the density becomes zero. Thus no sound is heard when a bell is struck inside a receiver from which the air has been exhausted as completely as possible.

[For the displacement of a particle which executes a simple harmonic vibration is

$$a \sin \frac{2\pi}{\mathrm{T}} t,$$

where T is the periodic time, and a is the amplitude. The speed of motion, which is the time-rate of variation of the displacement, is therefore

$$\frac{2\pi}{\mathrm{T}} a \cos \frac{2\pi}{\mathrm{T}} t,$$

and the energy of the particle is

$$m \frac{2\pi^2}{\mathrm{T}^2} a^2 \cos^2 \frac{2\pi}{\mathrm{T}} t,$$

where m is the mass. In an interval of time τ, which is very large in comparison with T, the energy may be regarded as constant and equal to its average value during the period τ. But this is

$$\frac{1}{\tau} \int_0^{\mathrm{T}} m \frac{2\pi^2}{\mathrm{T}^2} a^2 \cos^2 \frac{2\pi}{\mathrm{T}} t \, . \, dt = \frac{1}{\tau} \frac{m\pi^2 a^2}{\mathrm{T}^2} \int_0^{\tau} (\cos \frac{4\pi}{\mathrm{T}} t + 1) dt$$

$$= \frac{1}{\tau} \frac{m\pi^2 a^2}{\mathrm{T}^2} \left[\frac{\mathrm{T}}{4} \sin \frac{4\pi}{\mathrm{T}} t + t \right]_0^{\tau}.$$

The first term vanishes since τ is very large in comparison with T, and hence the kinetic energy is $m\pi^2a^2/T^2$ per particle, or $\rho\pi^2a^2/T^2$ per unit volume.

The maximum value of the kinetic energy per particle is $2m\pi^2a^2/T^2$, and so the largest possible value of the kinetic energy per unit volume is $2\rho\pi^2a^2/T^2$. Hence the total energy per unit volume of a system of particles in simple harmonic vibration is (in an interval of time which is large in comparison with the period of a complete vibration) one-half kinetic and one-half potential. This statement will obviously still be true, without the above restriction regarding the interval of time, provided that the number of particles per unit volume is very great, and that the wave-length of the disturbance is very small.]

In still homogeneous air sound spreads outwards from the source uniformly in all directions; and therefore particles, whose vibrations are in a given phase, lie on the surface of a sphere, the radius of which grows uniformly. Since the energy of vibration is distributed over the surface of this growing sphere, it follows that the intensity diminishes proportionately as the surface increases. It is therefore inversely proportional to the square of the distance from the source.

Sound travels faster than usual when the air through which it moves is blowing in the direction in which the sound moves, for the two speeds are simply superposed. Similarly, when the wind is blowing in a direction opposite to that in which the sound moves,

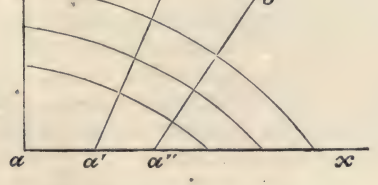


FIG. 76.

the speed is distinctly diminished. And, in addition to this, motion of the medium affects the distance at which the sound may be audible. For, if ab represent the (vertical) front of a plane-wave of sound travelling with the wind in the direction ax, it is evident, since the motion of the upper strata of air is less retarded by

friction than that of the lower strata, that the wave-front gradually becomes more and more inclined to the vertical as it moves forward. Successive new positions are indicated by the lines $a'b'$, etc. And the sound, instead of travelling straight forwards, is thrown down towards the ground in the manner indicated by the curved lines in Fig. 76. When the wind blows in a direction opposite to the direction of motion of the sound, the sound is thrown up from the earth's surface so as to be inaudible at comparatively short distances.

162. *Reflection of Sound.*—When sound strikes an obstacle, it is reflected in such a way that *the reflected ray* (the word is used by analogy from the phenomena of light, which, we shall see subsequently, are due also to wave-propagation) *and the incident ray*

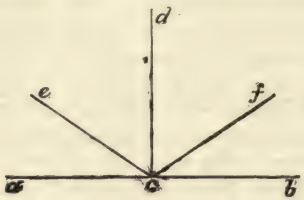


FIG. 77.

make equal angles with, and lie in a plane passing through, the normal to the surface. Thus, if *ab* (Fig. 77) represent a plane surface from which the ray *ec* is reflected in the direction *cf*, the line *cf* lies in a plane passing through *ec* and the normal *cd*, and the angles *ecd* and *fcd* are equal.

This law is identical with that which obtains in the reflection of light, so that all the results which are deduced in Chap. XVI. regarding the reflection of light from surfaces will at once apply to the case of reflection of sound. The law can be deduced as a result of the fact that sound consists in wave-motion, in precisely the same way as that in which the corresponding law is deduced in § 186 as a result of the wave-theory of light.

Echoes are due to the reflection of sound from buildings, rocks, trees, clouds, etc. They occur even when there is no visible object to account for their existence. In this case the reflection must occur at the common surface of two large masses of air of different densities, or containing very different amounts of moisture per unit volume. If the reflecting surface be curved, the reflected sound may be conveyed to a focus so as to be much more distinctly heard than the direct sound.

163. *Refraction of Sound.* — Since the speed of sound has different values in media of different densities, and since, in any one homogeneous medium, the sound spreads out uniformly in all directions from a centre of disturbance, it follows that the direction in which a ray is travelling is, in general, suddenly changed when the sound passes from one such medium into another. If it pass from a less dense into a more dense medium, the direction of propagation is inclined to the normal at a smaller angle after the interface is passed than before; and the opposite statement holds when the first medium is denser than the second. This phenomenon is known as the refraction of sound. The law is that *the incident*

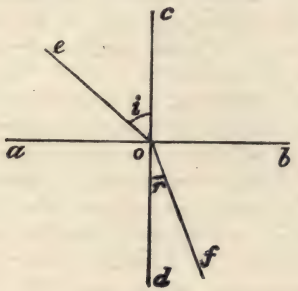


FIG. 78.

and the refracted rays lie in one plane with the normal to the refracting surface, and make with it angles whose sines bear a constant ratio to each other. If *ab* (Fig. 78) represent the intersection of the common surface of the two media by the plane of the paper, while *cd* is the normal, and if *eo* be the direction in which the sound impinges upon the surface, while *of* is the direction which it takes after entering the second medium, the angles $i = eoc$ and $r = fod$ are connected by the relation $\sin i = \mu \sin r$, where μ is a constant.

These results are identical with those which are observed in the refraction of light (§ 187), and the reasoning of § 200 may be applied directly to the present case.

A lenticular bag, filled with carbonic acid gas, has been found to convey sound to a focus in precisely the same manner that a glass lens conveys light to a focus.

164. *Diffraction of Sound.*—When sound enters a room by an aperture such as a window, it is equally well heard at all parts inside the room. It bends round so as to penetrate every portion, and does not cast a sharp 'shadow' of an obstacle as light does. This bending

of sound into the region behind an obstacle is termed *diffraction.* We shall see afterwards that the same phenomenon also occurs, under suitable conditions, in the case of light.

But it is possible, by proper means, to produce sound shadows. The necessary condition is that the aperture through which the sound passes, or the obstacle by which it is intercepted, shall be large in comparison with the wave-length of the sound. Thus a sound made at one side of a steep high bank may be totally unheard at the other. So, also, a sound may be clearly heard through a hollow between two mountains, while it is inaudible in the region behind either mountain. And, indeed, by using sound of sufficiently short wave-length, we can produce comparatively sharp sound shadows of obstacles which are only a few inches in diameter. (For an explanation of the phenomenon, see the discussion of the diffraction of light, Chap. XVIII.)

165. *Interference of Sound.*—As sound consists physically of waves of condensation and rarefaction, it follows that two sounds may 'interfere' with each other, according to the usual laws of interference of waves. Thus, if two sounds are travelling in the same direction through a given medium, and if the intensity and wave-length of these sounds are identical, there will be no resultant disturbance of the medium, *i.e.*, no sound will be heard, provided that the maximum condensation due to one of the disturbances occurs simultaneously with the maximum rarefaction due to the other. If both condensations, and therefore both rarefactions, take place together, a sound of four times the intensity will be heard, for the resultant amplitude of vibration of the particles of the medium is doubled. And, if the wave-lengths of the separate disturbances are not precisely identical, the resultant sound will periodically vary in intensity from the former (zero) to the latter (quadruple) value.

166. *Pitch.*—The resultant vibration of a particle of air is, in general, extremely complex; but, when a pure tone is transmitted through air, the vibration of each particle is simple harmonic. There can, therefore, be no difference between one tone and another except such as is due to a difference in the amplitude of vibration or to a difference in the period of vibration. As the amplitude of vibration determines the intensity of the sound, we infer that the pitch of a note depends upon the frequency of vibration, *i.e.*, upon the number of vibrations which are executed per second.

That this is actually the case may be roughly ascertained by means of very simple apparatus. If a piece of cardboard be pressed against the edge of a toothed wheel, which revolves at a

definite rate, a sound is emitted which is of a fairly definite pitch. As the speed of the wheel increases, *i.e.*, as the number of impacts per second between the teeth of the wheel and the cardboard increases, the pitch of the note rises; and, to a note of given pitch, there corresponds a definite rate of rotation of the wheel, and consequently a definite rate of vibration of the cardboard.

A much more accurate proof is obtained by means of the *syren*. This instrument consists essentially of a perforated disc of metal, the perforations being arranged in a circle round the centre, as in the figure. A tube, through which air is driven, is placed behind the disc; and, as the disc revolves, each successive opening in it

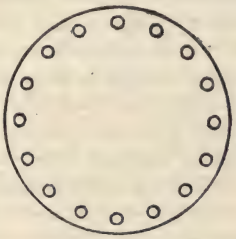


FIG. 79.

comes opposite the end of the tube. As each blast of air passes through, a state of condensation is produced, which is succeeded by a rarefaction in the interval between two blasts. Hence a sound-wave of constant period is set up when the disc revolves at a uniform rate. The rate of rotation, and the number of perforations in the complete circle, being known, we get at once the number of vibrations per second corresponding to any note of given pitch.

When the disc revolves very slowly, no musical sound is heard, but each separate air-blast can be distinctly heard. As the speed increases, the ear ceases to distinguish the separate pulses, and a note of very low pitch becomes audible. The speed still increasing, the pitch of the note produced becomes higher and higher, and at last it becomes so high that the note is no longer audible. The limits of audibility vary considerably in different observers, but, roughly speaking, a sound becomes inaudible as a note when the rate of vibration falls short of 20 times or exceeds 70,000 times per second. The fact that a melody is perfectly heard at different distances from the source shows that the speed of sound does not depend upon the wave-length.

The apparent pitch of a note depends upon the relative motion of

the hearer and the instrument upon which the note is sounded. If the hearer approaches the instrument, the pitch of the note seems to be heightened, because more than the normal number of vibrations reach his ear in a given time; and, conversely, if he recede from the instrument, the pitch of the note will appear to be lowered. As a particular case, if the rate of recession were greater than the speed of sound, no sound could be heard.

167. *Musical Intervals.*—If any particular note be taken as a keynote, it is found that there are certain other notes, definitely related to it as regards pitch, which produce specially pleasing effects upon the ear when sounded with each other or with the keynote. These are, therefore, the notes which are employed in the ordinary major and minor scales, and the difference in pitch of any two notes is called the *interval* between them. When two notes have the same pitch, they are said to be in *unison.*

The chief interval is the *octave*, and it is found that, in order to produce the octave of any given note, the rate of vibration must be exactly doubled. The intervals into which the octave is subdivided are not equal. The following table indicates their values. The first column gives the names of the intervals which separate each note from the keynote, and the relative rates of vibration in the two notes forming each interval are given in the second column. For example, nine vibrations are performed in the higher of two notes separated by the interval of a second in the time in which eight vibrations are performed in the lower:

Unison	$\frac{1}{1}$.
Second	$\frac{9}{8}$.
Minor Third	$\frac{6}{5}$.
Major Third	$\frac{5}{4}$.
Fourth	$\frac{4}{3}$.
Fifth	$\frac{3}{2}$ or $2(\frac{3}{4})$.
Minor Sixth	$\frac{8}{5}$ or $2(\frac{4}{5})$.
Major Sixth	$\frac{5}{3}$ or $2(\frac{5}{6})$.
Minor Seventh	$\frac{16}{9}$ or $2(\frac{8}{9})$.
Major Seventh	$\frac{15}{8}$.
Octave	$\frac{2}{1}$.

A very little consideration will show that in order to find the sum of two intervals we must multiply together the fractions given above for each interval. Thus a fifth is the sum of a minor third and a major third. Also, to find the difference between two intervals we must divide the fraction corresponding to the larger by

the fraction corresponding to the smaller. The second method of writing the fractions corresponding to the minor seventh, the major and minor sixths, and the fifth, shows that the interval between the notes so indicated and the octave are respectively a second, a minor third, a major third, and a fourth.

168. *Vibrations of Rods.*—We shall assume that the extent of the vibrations is such that Hooke's Law (§ 135) is followed. In this case, since the period of vibration is independent of the extent of vibration, a musical note of constant pitch will be heard, provided that the rate of vibration is sufficiently rapid. Two kinds of vibrations have to be considered.

Transverse Vibrations.—We see from §§ 51, 63 that the time of simple harmonic vibration of any material system varies directly as the square root of the mass to be moved and inversely as the square root of the stress called into play by a given displacement. It therefore, in the case under consideration, varies inversely as the square root of the flexural rigidity. Now both the flexural rigidity and the mass of a rod are (§ 133) proportional to its breadth, and so *the period of transverse vibration is independent of the breadth of the rod.* Again, the mass of the rod is proportional to the length, while the rigidity varies inversely as the cube of the length. Hence the period is *proportional to the square of the length.* Similarly, *it varies inversely as the thickness*, since the rigidity is directly proportional to the cube of the thickness.

Hence we conclude that the time of transverse vibration of similar rectangular rods is in direct proportion to their linear dimensions. And considerations of a like kind enable us at once to extend this statement to the case of similar rods of any form; for the masses of such rods are in proportion to the cubes of their linear dimensions, while their rigidities vary as the first power of the dimensions.

Longitudinal Vibrations.—If a rod be fixed at one end and receive a smart blow on the other end in a direction parallel to its length, a wave of compression will travel along it to the fixed end, from which, by reflection, it will return to the free end. The rod will now extend, because of its elasticity, to a length greater than its normal length, and a wave of rarefaction will travel along it to the fixed end at which it too will suffer reflection. The free end is obviously a loop, or place of maximum motion, while the fixed end is a node. Hence the length of a wave is four times the length of the rod, and the time occupied by the disturbance in travelling from end to end of the rod is therefore one-quarter of the periodic time of longitudinal vibration.

If the rod be fixed at both ends, the wave-length is equal to twice the length of the rod, since each end is now a node.

In each case the periodic time is obtained by dividing the length of the wave by the speed of the disturbance. The speed is given by equation (7), § 160, k being in this case the constant called Young's modulus. From this we see that the period is proportional to the length of the rod and the square root of the density, and is inversely proportional to the value of Young's modulus, while it is independent of the sectional area.

The rod which is fixed at both ends has twice the rate of vibration of a similar rod fixed at one end only. A rod which is free at both ends, and has its node at the centre, obviously vibrates at the same rate as does a like rod with both ends fixed.

169. A rod which is fixed at one or both ends may have more than one mode of transverse or longitudinal vibration. In Fig. 80 a represents the fundamental mode of vibration of a rod fixed at one end, while b and c represent the modes which stand next to it in order of simplicity. This is the case of a vibrating tuning-fork. When the fork is feebly excited, the fundamental mode of vibration is the chief one which occurs; but, under violent excitation, the forms b, c, and other higher forms, are superposed upon it. The extra node in b is distant from the fixed end two-thirds of the whole

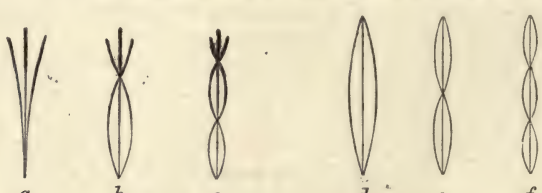


FIG. 80.

length of the rod, since the free part and the looped part must have a common period of vibration. The length of the free part in b is therefore one-third of the length of the free part in a, and consequently the period of vibration in the case b is one-ninth of the fundamental period. A reference to the table of § 167 will therefore show that the note given out by a fork vibrating in the mode b differs in pitch from the fundamental note by the interval of three octaves and a second.

In the figure, d, e, and f, show the three simplest modes of transverse vibration of a rod which is fixed at both ends. The interval between the notes given out by similar rods vibrating as in d and e

is two octaves, while the interval in the cases indicated by d and f is the same as between those indicated by a and b.

The modes of longitudinal vibration of rods are precisely identical with those of air-columns, which will be discussed shortly.

170. *Vibration of Plates.*—A plate may be set in vibration by drawing a bow across its edge. The modes of vibration depend on the form of the plate, and may be greatly varied by forcing certain points to lie upon nodal lines—which may be done by keeping the fingers in contact with the desired points.

As in § 168, we can without difficulty deduce the result that the periods of vibration of similar plates (similar, of course, as regards thickness as well as form) of the same material are in proportion to their linear dimensions. Also, as in the section referred to, we may show that the period is inversely as the thickness of a plate of given form and area. By combining these two results we see that the periods of vibration of similar plates of the same thickness are proportional to the squares of their linear dimension, that is, to their areas.

171. *Vibrations of Strings.*—The speed, v, with which a disturbance runs along a stretched cord is (§ 73) proportional to the square root of the tension, T, and is inversely proportional to the square root of μ, the mass per unit length of the string. Hence, if λ be the wave-length, the time of a complete vibration is

$$\frac{\lambda}{v} = \frac{\lambda\sqrt{\mu}}{\sqrt{\mathrm{T}}} = \frac{\lambda\sqrt{\rho s}}{\sqrt{\mathrm{T}}},$$

where ρ is the density of the material of the string and s is its area of cross-section.

The figure below shows the three simplest forms of vibration of a stretched string which is fixed at its ends. Any disturbance of the

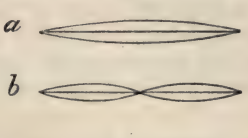


FIG. 81.

string is propagated with equal speed in both directions along the string from the point of disturbance, each part being reflected when it arrives at a fixed end. Hence, (§ 53), the string is thrown into

segments, the wave-length in each case being equal to twice the length of a loop.

Thus the above equation enables us to assert that *the fundamental period of vibration of a string is directly proportional to its length, to the square root of its density, and to the square root of its sectional area, and is inversely proportional to the square root of its tension.*

If a string of length l vibrates n times per second (determined by means of the syren), the speed of the wave is $2ln$.

The higher vibrations correspond to tones which are respectively one, one-and-a-half, two, two-and-a-quarter, etc., octaves above the fundamental tone. It is interesting to compare this result with the corresponding result in the case of a vibrating rod which is fixed at both ends. In the latter case, the transversely propagating force is the flexural stress; in the present case, it is the tension of the cord, in which flexural stress has no existence.

172. *Vibration of Air-Columns.*—When a condensation reaches the closed end of a pipe, or tube, containing air through which sound is travelling, the increase of pressure which takes place can only be relieved by expansion of the air backwards along the pipe. Similarly, when a rarefaction reaches the closed end, the consequent diminution of pressure can only be changed by a flow of the air contained in the pipe towards that end. In other words, the wave is simply reflected at the closed end, the result being that two sound-waves are travelling simultaneously along the pipe, in opposite directions, with equal speed. And the state of pressure due to each is always in the same phase at the closed end of the pipe. But, at a certain distance from the closed end, an outgoing condensation meets an incoming rarefaction. Hence, on the assumption that no energy has been lost in the process of reflection, the pressure has here its normal value. And this normal condition is maintained constantly so long as the wave-length of the disturbance is unaltered, for the two oppositely-travelling waves are always in opposite phases as regards pressure at this place, which is evidently distant by one-quarter of a wave-length from the closed end of the pipe.

The conditions of motion of the particles of air are simply reversed when the wave is reflected, and, consequently, the incident and the reflected disturbances are in opposite phases as regards vibration when they are in the same phase as regards pressure or density. Hence the places of maximum variation of pressure are places of no vibration, and the places of uniformly normal pressure are those of maximum vibration. Thus a node occurs at the closed

end of the pipe, and other nodes appear at equal intervals of half a wave-length.

It is obvious that a loop, or place of greatest motion, must occur at the open end of a pipe. For, when a condensation reaches it, the air is free to expand outwards—much more free to do so, in fact, than to expand inwards—and consequently the condensation is succeeded by a rarefaction which is propagated back through the pipe. Similarly, an incident rarefaction is filled up by the influx of the surrounding air, and so a condensation is returned. Thus we see that there is great vibration at the open end of the pipe, which therefore corresponds to a loop. [This might otherwise have been determined by means of the consideration that, in the immediate neighbourhood of the open end, the air-pressure is uniform and equal to the normal atmospheric pressure. Consequently the open end is a place of maximum vibration.]

Stopped and Open Pipes.—A pipe which is closed at one end is called a 'stopped' pipe, while one which is open at both ends is called an 'open' pipe.

In a stopped pipe the nodes evidently occur at the same positions as do the nodes of a transversely-vibrating rod which is fixed at one end. This is indicated in the figures below, in which the vertical distance between the waved lines at any part of the length of the pipe may be supposed to indicate the extent of the vibrations at that part. The first figure shows the fundamental mode of vibration. The wave-lengths, and consequently the periods of vibration, in all the possible modes are evidently inversely as the odd numbers 1, 3,

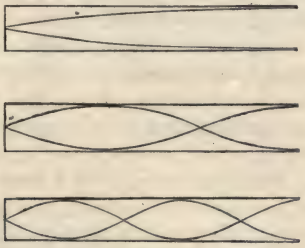


FIG. 82.

5, etc.; and, in the fundamental mode of vibration, the length of the wave is four times the length of the pipe; for a condensation which leaves the open end has its phase reversed when it again reaches that end and is reflected. Two reversals are therefore necessary in order that the original phase may be attained.

In an open pipe the nodes occur at the same positions as do the nodes of a transversely-vibrating rod which is not fixed at either end. The positions are indicated below. The possible periods of vibration are inversely as the natural numbers 1, 2, 3, etc. The

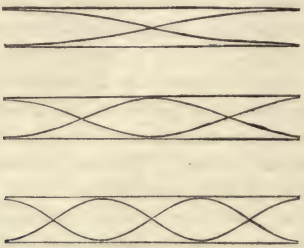


FIG. 83.

wave-length in the fundamental mode of vibration is twice the length of the pipe; for a condensation passing from one end of the pipe is reflected from the other end as a rarefaction, and leaves the original end once more as a condensation. Hence an open organ-pipe which has the same fundamental tone as a given closed organ-pipe must be twice as long as the closed one is.

Speed of Sound.—In order to determine very accurately the speed of sound in any given gas we merely require to sound an organ-pipe of given length which is filled with that gas. By means of the syren the number of vibrations which are made per second is found. And if l be the length of the pipe, while n is the number of vibrations made per second, the speed is $4nl$, or $2nl$, according as the pipe is closed or open. Under ordinary atmospheric conditions the speed of sound in air is about 1,120 feet per second.

173. *Partial Tones.—Resonance.*—From the results of the last few sections it is now evident that no note given out by a musical instrument is, strictly speaking, a pure tone; in most cases it very obviously is not so. Thus the 'tongue' of a reed is simply a rod or strip of metal which is fastened at one end, and is caused to vibrate by means of a current of air which is rendered intermittent by the vibrations which it produces in the tongue; and these air-pulses produce a note in which occur tones corresponding to the various forms of vibration which the tongue assumes.

These constituent tones of a note are called the *partial-tones;* and it is usual to distinguish between the *fundamental tones* (corresponding to the fundamental mode of vibration of the sounding body) and the higher partial-tones, which are termed the *overtones.*

Much greater force is required to set a body in rapid vibration than to set it in slow vibration, and consequently the overtones of a note are feebler than the fundamental tone—each overtone, taken in order of pitch, being weaker than the preceding one. In this way a note is said to have the same pitch as its fundamental tone.

The unaided ear is able to detect in great measure the various overtones which are present in a given note; but, in this analysis, it may be greatly aided by means of instruments, the action of which depends upon the principle of *resonance*, according to which *any sounding body can readily absorb, and give out again of itself, a sound which is emitted by another sounding body which has a period of vibration identical with its own.*

To understand this principle, we need only refer to a well-known dynamical analogy: A pendulum of given length has a definite period of vibration, and oscillations of great magnitude may be induced in it by the application to the bob of the *feeblest* impulses, provided only that these impulses are communicated regularly at instants the interval between which is equal to the natural period of oscillation of the pendulum. The effects of all the impulses are in the same direction, and so the total effect may be very large although each single effect is excessively small. And, similarly, the feeble periodic impulses which are communicated through the medium of the air to a body which is capable of giving out sound may, by being properly timed, set that body in such a state of vibration as to give out the sound of itself after the original note has ceased.

The *resonator* of v. Helmholtz consists of a hollow brass ball with two apertures at opposite ends of a diameter. (This is indicated in Fig. 84.) Sound is communicated to the air in the ball through the large aperture, and the small aperture is applied to the

FIG. 84.

ear. The air inside the ball has a definite fundamental period of vibration, and therefore that tone (should such exist) which corresponds to this period is very distinctly heard—all others being entirely, or almost entirely, suppressed.

In many cases it is desirable to intensify the fundamental tone of a note. The principle of resonance shows that this may be done by associating with the sounding body a resonating column of air which has the same fundamental period of vibration. Thus a reed is fitted at one end of an open pipe of the proper length; and a tuning-fork is attached to a 'sounding-box,' which is simply a closed pipe of such a length that the contained air-column has the same fundamental period of vibration as the tuning-fork has. The box itself is set in vibration by the fork, and so the motion is communicated to the enclosed air. The cavity of a violin acts similarly as a resonator.

The sounding-board of a pianoforte is thrown into vibration by the wires which are attached to it, and so increases the mass of air which is made to vibrate when a note is struck on the instrument. Thus, by its resonance, it intensifies the sound.

174. *Quality.*—There is between two pure tones of the same pitch no difference of quality such as that which distinguishes the same note when played on two different instruments. We therefore conclude that difference of quality between two notes is due to the existence of partial-tones.

Experiment shows that this is actually the case, and it indicates further that the quality of a note depends upon the number of partial-tones which are present in it; that, the number being the same, it depends upon the particular set of tones which go to make up that number; and, lastly, that it depends upon the relative intensities of the partial-tones.

[It is independent of the particular phase in which any one of the constituent vibrations may be when the phases of the others are given. Now the nature of the resultant vibration essentially depends upon this condition (§ 52, (1)), and so the above result appears somewhat startling. Its truth is due simply to the fact that the human ear is an instrument which resolves a compound vibration into its separate constituents, and thus mere alteration of phase of a constituent has no effect upon the nature of the sound which is heard.]

As a special example, we may instance the difference of quality between two notes of the same pitch sounded respectively on a closed and an open organ-pipe. In a closed pipe the odd partial-tones alone occur, while in the open pipe both the odd and the even tones are present.

175. *Beats.—Consonance and Dissonance.*—It has already been pointed out (§ 165) that the intensity of the resultant sound due to two component vibrations of slightly different period varies

periodically from a minimum to a maximum. These regularly-occurring maxima constitute *beats*. One beat occurs in the time in which one vibration gains a complete period upon the other; and, consequently, the number of beats which occur per second is equal to the difference of the number of vibrations which take place per second in the two component tones.

When the beats succeed each other with too great rapidity, the ear becomes unable to distinguish them from one another, but it still recognises a distinct discontinuity in the sound, which produces a harsh effect known as *dissonance*.

The number of beats which occur per second when two given pure tones are sounded depends, of course, upon the absolute, as well as upon the relative, pitch of these tones. When the tones are in the neighbourhood of the middle C, maximum dissonance is produced when they differ in pitch by about half a tone; and there is almost no dissonance when the interval is a minor third.

On the other hand, beating may occur between the partial-tones of two notes, and the result is that there may be considerable dissonance between the two, even although their fundamental tones differ by more than a minor third. Indeed, from this cause, none of the intervals below the octave form a perfect concord. Still, it is only in the cases of the second, and the major and minor sevenths, that the result is actually classed as a discord. In some cases the greatest dissonance, as above defined, does not produce the most disagreeable effect upon the ear.

176. *Combination Tones.—Dissonance of Pure Tones.*—When the rapidity of the beats is sufficiently great, the effect upon the ear is that of a musical note the pitch of which is the same as that of a tone in which the number of vibrations per second is equal to the difference (or sum) of the number of vibrations per second in the two tones which are producing the beats. Such tones are called *combination tones*.

Combination tones of a higher order may be produced between the first combination tone and either of the primary tones, and so on.

Beating may occur between a combination tone and a primary, or between two combination tones. The result is that dissonance may occur in the case of two pure tones where the interval is greater than a minor third. Thus, when the interval is a major seventh, the higher tone will make 450 vibrations per second if the lower makes 240 per second. The combination tone has therefore 210 vibrations per second, which produces 30 beats per second with the lower primary.

CHAPTER XV.

LIGHT: INTENSITY, SPEED, THEORIES.

177. *Rectilinear Propagation. Intensity.* — In a homogeneous medium, light, in general, moves in straight lines. The fact that the shadow of an object which is cast on the ground by sunlight is not defined with mathematical accuracy, but has a more or less blurred edge, does not disprove the statement. The indistinctness of the boundaries of the shadow is due to the finite size of the sun's disc, the light proceeding from each point of which produces a separate shadow.

The cases in which the above rule is departed from will be discussed in Chap. XVIII.

The total intensity of a luminous source is measured by the amount of light which it emits per unit time, and the intensity of the light at any given point of the medium is measured by the quantity which falls per unit time upon unit area taken perpendicular to the direction in which the light moves at that point. And, since light moves out uniformly in all directions from a point-source in a homogeneous medium, it follows that its intensity varies inversely as the square of the distance from the source; for the same total quantity is, at different instants, spread over the surfaces of different concentric spheres, the areas of which vary directly as the squares of their radii. (We assume, of course, that the medium is one which does not absorb the light in its passage through it. If any absorption did occur, the quantities of light passing through the different concentric spheres could not be equal.)

Instruments used for the purpose of comparing the intensities of different sources of light are called *photometers*. The simplest form of photometer consists of a sheet of paper upon which there is a grease spot. If the paper be illuminated from behind, the spot appears bright; if it be illuminated from the front, the spot seems dark; if it be illuminated to an equal extent on both sides, the spot vanishes. Under the latter condition the intensities of the sources are inversely as the squares of their distances from the spot.

In another form of the instrument two grease spots are employed, each being illuminated from behind by one source alone. The distances of the sources are varied until the two spots appear equally bright.

178. *Speed.*—In last section we have spoken of the *motion* of light. The use of the term is justified by the fact that a flash of light is not simultaneously seen by two observers who are situated at different distances from the source. When the distance between the two points of observation is not very large—a few miles, say—the interval of time which is occupied by the light in passing from one point to the other is so small that it cannot be measured except by very special means. But there are two astronomical methods of determining the speed of light which do not involve the measurement of a small interval of time.

The first of these is due to Römer, who observed that the eclipses of Jupiter's satellites do not appear to recur at equal intervals of time, and pointed out that this would be a necessary consequence of the finite speed of light.

In order to understand more clearly how this may be, we may take an illustration from the phenomena of sound. If an observer be situated at a fixed distance from a point at which a gun is fired off at equal intervals of one minute, he will hear the report at equal intervals of one minute. But if, between two successive discharges of the gun, he move nearer to it, he will hear the next report at a shorter interval of time than one minute; while, if he move farther from it, the interval will necessarily be greater than one minute. (When applied to light (§ 204), this principle is usually called Döppler's principle.) The speed of sound may be determined from the results of two such observations. Let τ be the time which elapses between two successive discharges, and let t be the interval noted by the observer between two successive reports when he has meanwhile increased his distance from the gun by the amount d. Then, t' being the (unknown) time taken by the sound to pass over the distance d, we get $t=\tau+t'$, and therefore the speed of sound is given by the quotient of d by $t-\tau$.

Now the eclipses of Jupiter's satellites occur at instants which are very accurately calculable from known astronomical data. But the observed instants at which the eclipses apparently take place as seen from the earth do not coincide with the calculated instants; and the errors at different times of the year are (assuming the finite speed of light) due to the variation of the distance between Jupiter and the earth. The greatest difference of apparent errors is the time which light takes to pass over the greatest difference of dis-

tance. But the greatest difference of distance between Jupiter and the earth is the diameter of the earth's orbit; and so we obtain the time taken by light to pass over this known distance.

The speed of light as deduced by this method is about 186,000 miles per second.

The other astronomical method is due to Bradley. He observed that the fixed stars appear to describe small ellipses on the surface of the heavens in the course of a revolution round the sun, each star being displaced from the centre of its elliptic path *in the direction of the earth's motion in its orbit, and each to the same amount.* He concluded that this was due to the finiteness of the speed of light as compared with the speed of the earth in its orbit.

A simple illustration may make this clear. On a still day, raindrops fall vertically downwards. But, if one moves forward with considerable speed, they do not seem to fall vertically; they apparently fall in a slanting line, which is inclined forwards from the vertical in the direction of the observer's motion. And it is evident that the apparent velocity of the drops is the resultant of their actual velocity and a velocity equal and opposite to that of the observer.

The light which comes from a star appears to come in a direction which depends in the same way upon the velocity of light and the velocity of the earth in its orbit. This latter velocity and the maximum angular displacement of a star from its true position being known, we can calculate the speed of light. The value which is obtained by this method agrees very closely with that obtained by Römer's method.

Fizeau was the first (1849) to determine the speed of light by direct experiment. He caused a beam of light to pass out through the gap between two of the teeth of a toothed wheel, the teeth and gaps of which were all of one size. This beam was reflected from a mirror, placed at a distance of a few miles from the wheel, in such a way that it passed back again through the same gap between the teeth of the wheel. The wheel was then caused to rotate, and, at a certain rate of rotation, it was found the light ceased to pass back between the teeth; the reason being that an adjacent tooth had moved into the place of the gap in the time that the light took to travel twice over the distance between the wheel and the mirror. The rate of rotation of the wheel, and the number of teeth which it contained, being known, the time which was taken by the light to pass over the given distance can be readily found. If N is the number of revolutions which it made per unit of time, while n is the total number of gaps and teeth in its circumference,

the speed of light is $2dNn$, d being the distance between the wheel and the mirror.

Fizeau's experiments were repeated some time afterwards by Cornu with improved apparatus; and, still more recently, a further improvement of the same method has been effected by Professor G. Forbes and Dr. J. Young.

In Foucault's method (recently improved by Michelson), a beam of light, after passing through a slit, falls on a mirror which can be made to rotate about an axis parallel to the slit. After reflection from the mirror, the light passes through a lens, which brings it to a focus on a fixed mirror. This mirror being so placed as to exactly reverse the course of the beam, the light once more falls upon the first mirror, and is reflected from it. If the latter is rotating, and has turned through a sensible angle in the time taken by the light to pass twice over the distance between it and the other, the beam will not pass back through the slit, but will be deflected from it through a measurable distance. The speed of light may be found in terms of the two distances just mentioned, together with the rate of rotation of the revolving mirror and the distance between it and the slit.

These two experimental methods give values of the speed of light which agree very closely with the values obtained by the two astronomical methods.

Foucault's method is so sensitive that it may be used successfully when the distance between the mirrors is only a few feet, and it lends itself readily to the determination of the speed of light in different media, such as glass, water, etc. The speed is found to be less in dense, than in rare, media.

179. *Theories.*—The transference of light involves motion of matter; for when light is absorbed by any body, increased motion of the particles of that body is generally produced. Indeed, in the radiometer (§ 153), visible motion of a considerable mass of matter may follow the absorption of light. Another marked example will be found in § 376.

Hence we see that transference of light implies transference of energy; and it is in this sense that we speak of light as a form of energy.

We are therefore limited to two suppositions regarding the physical nature of light. It may consist in the actual propagation of material particles, or corpuscles, from the luminous object; or, it may consist in the propagation of wave-motion through a material medium which fills space.

The former theory is known as the *Corpuscular Theory*, the latter as the *Wave Theory*, or *Undulatory Theory*, of light.

If the corpuscular theory were true, the mass of a corpuscle must be excessively small. For vision, according to this theory, is due to the impact of the corpuscles upon the retina; and the speed of these corpuscles is so great that, unless their individual mass were almost vanishingly small, the structure of the eye would be completely destroyed by the impact. The theory is met by a number of difficulties at the very outset. Thus it is somewhat difficult to account for the fact that the corpuscles have the same speed whatever be the temperature of the object from which they are projected. Again, the mass of a luminous body must be appreciably affected by the emission of particles; but there is no evidence of any such effect. Still, if we boldly overlook any such preliminary difficulties, we shall find that the theory enables us to account readily for many of the phenomena of light, although ultimately it fails us altogether.

On the wave-theory, vision is due to the communication of the vibrations of the assumed luminiferous medium (called the ether) to the nerve-ends of the retina. The molecules of a luminous body are (§ 202) in rapid vibratory motion, and this motion is communicated to the particles of the ether, and is propagated through it from particle to particle giving rise to a series of waves which travel with the speed of light. The investigation of § 161 has a direct application to the present case, and shows that the intensity of light is proportional to the square of the amplitude of vibration of the particles of the medium, and that the energy of the medium, when light passes through it, is one-half kinetic, one-half potential. (On the corpuscular theory the intensity must be proportional to the space-density of the corpuscles.) We shall find subsequently (Chap. XIX.) that the direction of vibration in the medium must be perpendicular to the direction of propagation of the waves.

The wave-length is the distance, measured in the direction of propagation, from any point to the next point where the motion is similar. (Compare § 157.)

The wave-theory is not without its difficulties—many of them, indeed, are of a most formidable nature. But, as will appear, the evidence in favour of it is of such an overwhelming nature that we now practically regard its truth as definitely established. Newton rejected it because he was unable to explain by it the rectilinear propagation of light. We now know that the existence of rays is a necessary consequence of the fundamental principles of the theory.

180. *Colour.*—Many different kinds of light are recognised. We speak of red light, blue light, etc. On the corpuscular theory the difference must be inherent in the corpuscles. On the wave-theory the difference is a mere difference of wave-length, or of vibrational period—which is only another way of stating the same thing, since the speed of propagation of light of all colours has, in free space, one definite value only.

CHAPTER XVI.

LIGHT: REFLECTION, REFRACTION, DISPERSION.

181. *Laws of Reflection.*—When a ray of light reaches the bounding surface of a homogeneous medium through which it is passing, it is, in part at least, bent back or reflected, and pursues a different, though still rectilinear, path.

The reflected and incident rays lie in one plane with, and make equal angles with, the normal to the surface.

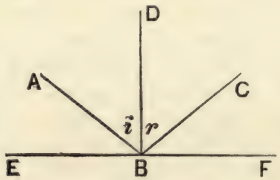


FIG. 85.

Let EBF (Fig. 85) represent a section of the bounding surface by the plane of the paper, and let BD be the normal to the surface at the point B whereon the incident ray AB falls. Then, BC being the reflected ray, the angles i and r, which AB and BC make with BD, are equal, and AB, BC, and BD all lie in one plane, which is normal to the reflecting surface. The angles i and r are called, respectively, the angle of incidence and the angle of refraction.

Many surfaces, such as those of chalk or of rough white paper, *scatter* the incident light in all directions. But this is merely a special case of reflection. At every point of such a surface rays are reflected in accordance with the above law; but the whole surface is practically made up of an excessively great number of very small planes, which are indiscriminately inclined in all possible ways.

The intensity of the reflected ray depends upon the angle of incidence, being greatest when the angle is large, and having its least

value when the incidence is perpendicular. It varies (the intensity of the incident ray being fixed) with the nature and the state of surface-polish of the reflecting substance, and it depends also upon the nature of the medium through which the light is travelling.

182. *Reflection from Plane Surfaces.*—If a ray of light be emitted from a point B (Fig. 86) and reach a point A, after reflection

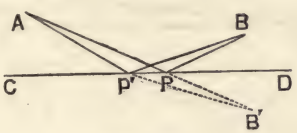


FIG. 86.

from a plane surface, CD, the actual length of the path APB is the shortest possible consistent with the condition of reflection at the given surface.

For an eye placed at the point A will see the light in the direction AP as if it came from a point B′, which is situated on the normal drawn from B to the surface. [This is so since the eye sees an object by means of a cone of rays: and the angle of the cone is unaltered by reflection since (Fig. 87) the angles which a_1p_1 and a_2p_2

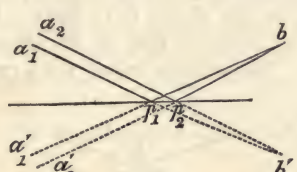


FIG. 87.

make with the reflecting plane are respectively equal to the angles which $p_1a'_1$ and $p_2a'_2$, the continuations of the lines bp_1 and bp_2, make with that plane. And so the continuations of a_1p_1 and a_2p_2 meet at a point b', which is such that bp_1b' and bp_2b' are both bisected by the surface; and therefore b' and b lie on the same normal to the surface, and are equally distant from it.] And any other path, AP′B, being equal to AP′B′, is greater than APB, which is equal to APB′.

The point B′ is called the *image* of the point B. If B were a body of finite size, each point of it would give rise to an image; and the whole congeries of these point-images constitutes the image of the body B.

183. *Reflection from Curved Surfaces.*—Let Q (Fig. 88) be the section of a spherical mirror by the plane of the paper. Let O be the centre of the sphere, and let a ray, UP, emitted by a luminous

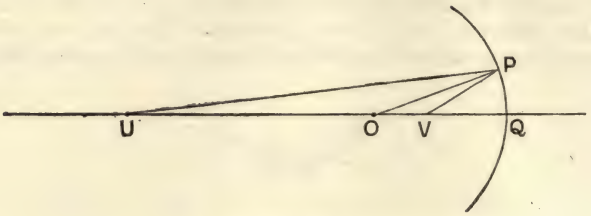


FIG. 88.

object at the point U, be reflected to the point V. We have PVQ = UPV + PUQ = 2UPO + PUQ. Therefore PVQ + PUQ = 2(UPO + PUQ) = 2POQ. When P and Q are nearly coincident this becomes approximately PQ/PV + PQ/PU = 2PQ/PO or 1/PV + 1/PU = 2/PO. If we denote by u, v, and r, the lengths of the lines PU, PV, and PO respectively, this gives

$$\frac{1}{u}+\frac{1}{v}=\frac{2}{r}.$$

This equation enables us to calculate the position of the image V when the position of the object U is given. If U be situated at infinity towards the left-hand side of the diagram, V is half-way between O and Q. This point is called the *principal focus* of the

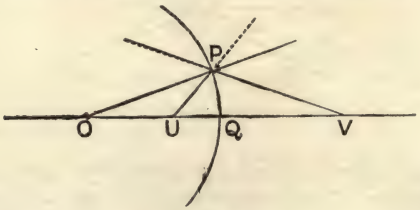


FIG. 89.

mirror. As U moves in from infinity V moves out to meet it, and the two points coincide at O, the centre of the sphere. The positions of U and V are now interchanged, and finally, when U is at the principal focus, V is at infinity towards the left. Whenever U comes nearer Q than the distance of a semi-radius, the quantity v becomes negative, that is, the image passes away (Fig. 89) behind the mirror, and gradually approaches it from infinity in this direction until both object and image coincide at Q.

In all cases V and U are interchangeable, that is to say, V may be the object and U will then be the image. A slight inspection of the two diagrams will make this clear.

The image, when on the same side of the mirror as the object is on, is called a *real* image; when on the opposite side, it is called a *virtual* image.

It is evident that an object on the convex side of the sphere can have a virtual image only. In this case the above formula becomes

$$\frac{1}{v} - \frac{1}{u} = \frac{2}{r},$$

which is the modification of the formula necessary to make it apply to the case of reflection from a convex spherical mirror.

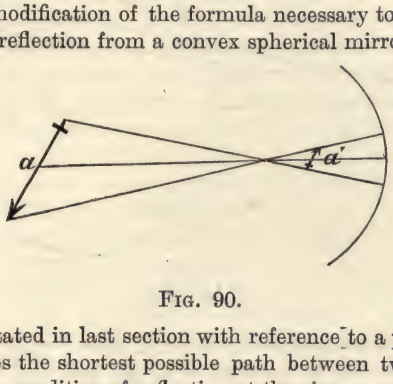


FIG. 90.

The law—stated in last section with reference to a plane mirror—that light takes the shortest possible path between two points consistent with the condition of reflection at the given surface, still holds in the case of *any* surface provided that we limit the statement to other paths which do not finitely differ from the actual path of the light. The necessity for this limitation will be evident if we consider

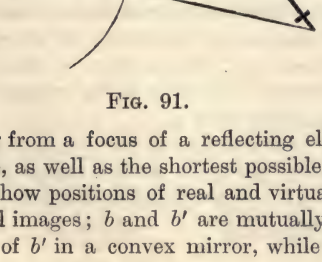


FIG. 91.

that light diverging from a focus of a reflecting ellipsoid may take the longest possible, as well as the shortest possible, path.

Figs. 90 and 91 show positions of real and virtual images: a and a' are mutually real images; b and b' are mutually virtual images; b being the image of b' in a convex mirror, while a, a', and b' are

the images of a', a, and b in a concave one. When formed by one reflection, or by an odd number of reflections, a real image is inverted, but a virtual image is not inverted. The positions of the various points of the image corresponding to given points of the object are found by means of the above formulæ.

184. *Caustics: Focal Lines.*—Let CBQ (Fig. 92) represent the section of a spherical mirror by the plane of the paper, and let PC,

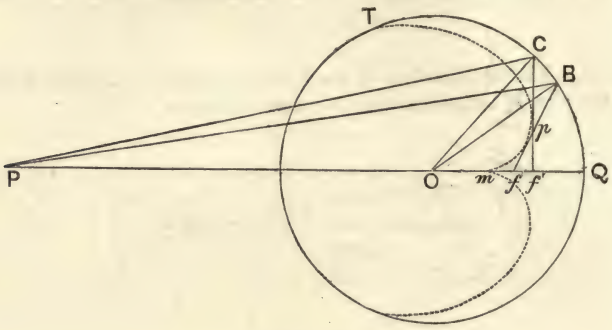


FIG. 92.

PB represent two rays which, diverging from the point P, fall upon the mirror: let also the reflected rays, Cpf' and Bpf, intersect in the point p.

Since the vertical angles at the intersection of Cp and OB (a radius) are equal, we have

$$\mathrm{OC}p+\mathrm{COB}=\mathrm{OB}p+\mathrm{C}p\mathrm{B}, \text{ or } \mathrm{OCP}+\mathrm{COB}=\mathrm{OBP}+\mathrm{C}p\mathrm{B}.$$

This gives $$\mathrm{COQ}-\mathrm{CPO}+\mathrm{COB}=\mathrm{BOQ}-\mathrm{BPO}+\mathrm{C}p\mathrm{B},$$

whence $$\mathrm{CPB}+\mathrm{C}p\mathrm{B}=2\mathrm{COB}.$$

[The results of last section follow as a particular case of this.]

Let CB be an infinitesimally small arc of constant length, and let T be the point at which the tangent from P meets the circle CBQ. CPB always diminishes, and therefore (by the above equation) CpB constantly increases as C moves from Q towards T. Hence the length of Cp always diminishes as its inclination to CO increases, until finally p coincides with T.

The locus of p is called the *caustic curve*, and is indicated by the dotted curve in the figure. It touches the circle at the point T and the line PQ at a point m, the position of which may be found by the formula of last section.

If the whole figure be rotated about the line PQ, the circle CBQ traces out the spherical reflecting surface, and the caustic curve traces a continuous surface which is called the caustic surface. All points on this surface are more intensely illuminated by the reflected light than any point which does not lie upon it, and the cusp (at m) is the place of most intense illumination. All rays which are reflected from the surface pass through the line PQ.

Let us suppose that a small, but finite, circular cone of rays falls upon the reflecting surface in the neighbourhood of the points B, C. All rays from points on the small circle, the pole of which is Q, and which passes through B, intersect PQ in the point f; and all rays from points on the similar circle through C pass through f'; and so on. It is evident, therefore, that a plane which is perpendicular to the axial line of the reflected cone, and which passes through the point in which the axial line intersects the line PQ, will cut the cone in an elongated figure-of-eight-shaped area, which may be regarded as a straight line, and is called the *secondary focal line.*

Again, a plane drawn perpendicular to the axial line through the point in which that line touches the caustic surface, cuts this surface in a circle, and all the reflected rays will pass through the plane in the immediate neighbourhood of a small, practically straight, portion of the circle; so that there is another, nearly linear, normal section of the reflected cone. This is called the *primary focal line.*

The two focal lines are mutually perpendicular.

An approximately circular section exists between the two linear sections. This is called the *circle of least confusion*, and is the place where the reflected light most nearly converges to a point.

185. The law of reflection follows readily from the principles of the corpuscular theory. Let pq (Fig. 93) represent the path of a

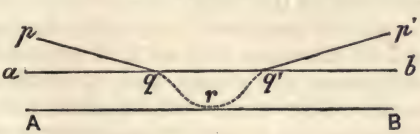


FIG. 93.

corpuscle, and let AB be the surface from which the corpuscle is reflected. When the corpuscle comes within a certain small distance from the surface, indicated by the line ab, it experiences the attraction of the medium which is bounded by the surface AB, and so is bent from its rectilinear path. The mutual action of the par-

ticle and the medium may alternate from attraction to repulsion many times according to an unknown law, but it must ultimately be a repulsion which (at r) stops the motion of the corpuscle towards the surface. The mutual forces still acting as before, the particle must now describe a path rq' precisely similar to rq, until, at q', being freed from the action of the reflecting medium, it describes a rectilinear path, $q'p'$, which is inclined to AB at the same angle as pq is. And, since the action of the medium is everywhere in lines perpendicular to the surface AB, the particle retains its velocity parallel to AB unaltered during the process of reflection, and the lines pq and $p'q'$ are in one plane with the normal to the surface.

186. The wave-theory also affords a ready explanation of the phenomena of reflection.

But, before dealing with this point, it is necessary to consider the explanation, on this theory, of the rectilinear propagation of light. Newton did not see how to account for it, and so supported the corpuscular theory. Huyghens was the first to show that the wave-theory furnishes a ready explanation of the phenomenon.

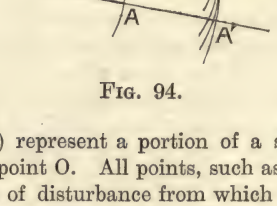


FIG. 94.

Let AB (Fig. 94) represent a portion of a spherical wave-front diverging from the point O. All points, such as a, b, c, on this surface become centres of disturbance from which secondary spherical waves diverge. With radius AA′, or BB′, equal to the distance which light will travel in a certain time, t, describe circles from a, b, c as centres. These circles will all touch another spherical surface, A′B′, concentric with AB. This constitutes the new wave-front. At all points of this surface secondary wavelets are superposed, and so a strong resultant effect may be produced. At no other points, besides those on A′B′, are the effects of the separate wavelets superposed, and the isolated wavelets are too feeble to produce light. Hence the rays included in the region AOB diverge outwards in straight lines.

The above explanation is due to Huyghens, and will suffice for

our present purpose. But, as Fresnel pointed out, Young's principle of interference is essential, in addition to the above, in order to make the demonstration rigorous. See § 226.

We shall assume for the sake of simplicity that we are dealing with a plane wave propagated through the luminiferous medium. Let ADF (Fig. 95) represent the reflecting surface, and let the given disturbance have reached the position ABC; so that ABC represents a portion of the plane wave-front, to which the rays (of which three are indicated in the figure) are everywhere perpendicular, the medium being assumed to be homogeneous and isotropic.

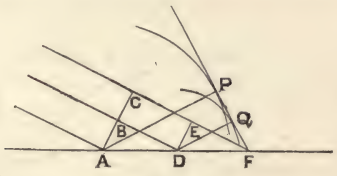


FIG. 95.

When the wave reaches the point A, the particles of the ether at that point are set in vibration and give rise to a spherical wave which spreads out from A as centre. If, from A, we draw a sphere with radius AP = CF, we get the position of this spherical wave when the disturbance originally at C has reached the reflecting surface. Similarly, DE being parallel to ABC, if we draw from D as centre a sphere with radius DQ = EF we get the corresponding position of the spherical wave which is originated at the point D when the wave-front reaches it. All such spheres touch a plane surface, PQF, which is, therefore, the wave-front after reflection. Thus we see that a plane wave remains a plane wave after reflection.

But, further, AP = CF, and AF is common to the two right-angled triangles ACF and APF. Therefore the angles CAF and PFA are equal, that is, the reflected wave-front has the same angle of inclination to the reflecting surface as the incident wave-front has. And this—since the rays are perpendicular to the wave-fronts—gives the known law of reflection.

187. *Laws of Refraction.*—A ray of light, on passing from one medium into another of different density, is in general bent from its original direction, and is said to be refracted. The angle which the refracted ray makes with the normal is called the angle of refraction.

The refracted and incident rays lie in one plane with the

normal to the surface and the sine of the angle of incidence bears a constant ratio to the sine of the angle of refraction.

The angles of incidence and refraction (Fig. 96) being denoted by the letters i and r respectively, the above law may be written

$$\sin i = \mu \sin r.$$

The constant, μ, is called the *index of refraction.* When the refraction takes place from a less dense into a more dense medium, μ is generally greater than unity; and, conversely, μ is usually less than unity when the light passes from a more dense into a less dense medium.

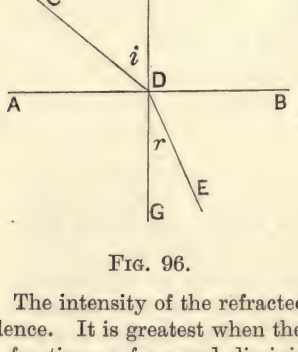


FIG. 96.

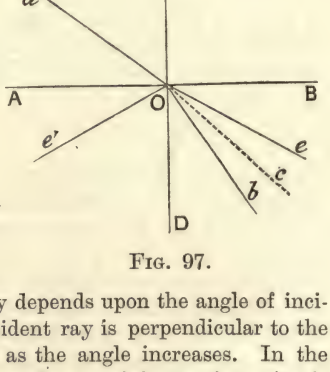


FIG. 97.

The intensity of the refracted ray depends upon the angle of incidence. It is greatest when the incident ray is perpendicular to the refracting surface, and diminishes as the angle increases. In the case of refraction into a denser medium, minimum intensity is attained at grazing incidence: in the case of refraction into a rarer medium, the minimum is reached when the angle of incidence is less than 90°; and at higher angles of incidence no refraction occurs.

188. *Refraction through a Plane Surface.*—Let a ray, aO (Fig. 97) fall upon the plane surface, AB, of a medium, the refractive index of which is μ—that of the first medium being taken as unity. If we assume μ to be greater than unity, the path of the ray in the second medium will be a line Ob, such that sin aOc = μ sin bOd, CD being perpendicular to AB at the point of incidence O. Also light travelling in the direction bO will emerge into the first medium in the direction Oa, which is such that

$$\sin b\mathrm{O}d = \frac{1}{\mu} \sin a\mathrm{O}c.$$

Now suppose aOC = 90°. The light will enter the second medium

in a direction Oc, such that $\sin cOD = \mu$; and, conversely, light travelling in the direction cO will emerge into the first medium so as just to graze along the surface. Light passing in the direction eO, where $eOD > cOD$, cannot enter the first medium at all, but will suffer total reflection in the direction oe', in accordance with the ordinary law. (This proves the concluding remark of last section.) The ray eo is said to suffer *Total Reflection*, and the limiting angle cOD is called the *Critical Angle*.

The refractive index of water, relatively to that of air, is about 4/3 for ordinary yellow light. Hence an eye placed underneath the surface of water will see objects above the surface by means of rays which are crowded into a cone, the sine of the semi-vertical angle of which is 3/4.

The angle of deviation of a ray from its original direction by a single refraction is $i - r$, i and r being respectively the angles of incidence and refraction.

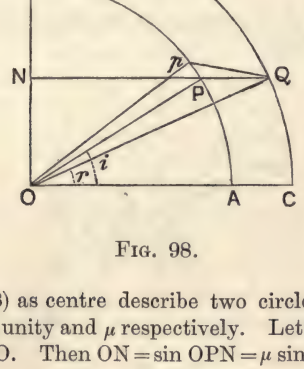


FIG. 98.

From O (Fig. 98) as centre describe two circles APB and CQD, with radii equal to unity and μ respectively. Let $COQ = r$, and draw QPN parallel to CO. Then $ON = \sin OPN = \mu \sin OQN = \mu \sin r$, and so $OPN = i$.

Now OP and OQ are fixed in length, and PQ is always parallel to OC; so that OPQ and OQP become more nearly right angles, and PQ becomes greater at an increasing rate as i increases up to 90°. Hence $i - r$ *increases faster and faster as either* i *or* r *increase uniformly.*

Next suppose that PQ moves out from OC through an infinitesimal distance into a position P′Q′, and let OQ′ be then turned back to coincide with OQ. P′ will then take a position p such that Ap is

greater than AP. By this process $i - r$ has increased by the amount Pp, and PQO $= r$ has increased by the amount pQP $=$ Pp cos i/PQ. Hence $r-(i-r)=2r-i$ has increased by the angle Pp (cos i/PQ -1). This vanishes when cos $i=$PQ. But cos $i=$NP, and hence the condition is NP $=$ PQ. This implies AC $<$ OA, that is, $\mu < 2$. Again, the increase of $2r-i$ is positive or negative according as NP $>$ or $<$ PQ. Hence it is always positive until the limiting angle is reached, after which PQ still increases, while NP diminishes, so that $2r-i$ diminishes continuously. In other words the condition NP $=$ PO or 2 cos $i=\mu$ cos r (since NQ $=\mu$ cos r) indicates a *maximum* value of $2r-i$.

This condition, combined with sin $i=\mu$ sin r, gives

$$3 \sin^2 i = 4 - \mu^2.$$

Similarly, the angle $3r-i$ increases under the same conditions by the amount Pp (2 cos i/PQ $-$ 1); and the condition for a stationary value is 2NP $=$ PQ or 3 cos $i=\mu$ cos r, which implies $\mu < 3$. As in the former case the condition indicates a maximum. From this along with sin $i=\mu$ sin r we get

$$8 \sin^2 i = 9 - \mu^2.$$

These results are of importance in the discussion of the primary and secondary rainbows.

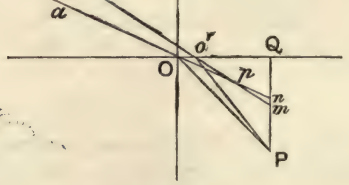


FIG. 99.

189. *Focal Lines: Caustics.*—Let OPo' (Fig. 99) represent an axial section of an infinitesimal circular cone of rays diverging from a point P placed in a dense medium. The section, by the plane of the paper, of the cone $a'pa$, by which the object P is actually seen by an eye situated in the rare medium, has its vertex at a different point, p, which is nearer the surface than P is.

Let i', r', and i, r, represent respectively the angles of incidence and refraction (regarded from the rarer medium) of the rays $a'o'$P

and αoP respectively. The angle of the cone, whose vertex is at P, is $r'-r$, and the angle of that whose vertex is at p is $i'-i$.

To find the relation between these angles we may write the equation $\sin i'=\mu \sin r'$ in the form $\sin(\overline{i'-i}+i)=\mu \sin(\overline{r'-r}+r)$. This gives $\sin(i'-i)\cos i+\cos(i'-i)\sin i=\mu\,[\sin(r'-r)\cos r+\cos(r'-r)\sin r]$, and so, remembering that $i'-i$ and $r'-r$ are vanishingly small, we get

$$\frac{i'-i}{r'-r}=\mu\frac{\cos r}{\cos i}=\frac{\tan i}{\tan r}.$$

But $i'-i=\mathrm{O}o'\cos i/\mathrm{O}p$, and $r'-r=\mathrm{O}o'\cos r/\mathrm{OP}$. Hence

$$\frac{\mathrm{OP}}{\mathrm{O}p}=\frac{\sin i}{\sin r}\cdot\frac{\cos^2 r}{\cos^2 i}.$$

Now let Op be prolonged to meet PQ (the line drawn through P perpendicular to the surface) in p', and we get $\mathrm{O}p'\sin i=\mathrm{OP}\sin r$;

so that
$$\mathrm{O}p=\mathrm{O}p'\frac{\cos^2 i}{\cos^2 r}.$$

This formula, which will be of use when we deal with spherica lenses (§ 194), shows at once (*cf.* § 184) that no more than two finitely distinct rays in any given vertical plane section can intersect in one point, p; but, when the cone is very small, all the rays in a given plane section pass approximately through one point.

We have next to consider the lateral divergence of the beam of light. This is obviously unaltered by refraction; for, after refraction, a ray remains in the same plane, passing through the normal, which contained it before refraction. Hence, laterally, the rays diverge from points on PQ. And thus, since the cone has some lateral thickness at the point p, we see that the emergent light appears to pass, at p, approximately through a small line, the length of which is perpendicular to PQ.

This is the primary focal line.

The secondary focal line is the intersection of the refracted cone with a plane drawn perpendicular to its axial line and passing through the point in which the axial line cuts the line PQ.

At perpendicular incidence, $\mathrm{O}p=\mathrm{O}p'=\mathrm{OP}/\mu$; so that the light appears to diverge from a point which is closer to the surface than the actual point in the ratio of 1 to μ. For example, the depth of water appears to be only three-fourths of its true depth.

The existence of the primary focal line implies the existence of a caustic surface. We may map out (Fig. 100) successive points,

p_1, p_2, etc., corresponding to successive positions of the point o, by means of the last equation above together with the condition $\sin i = \mu \sin r$. We thus obtain the caustic curve $p_0 p_1 p_2 o'$, which

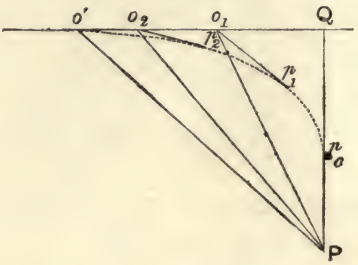


FIG. 100.

touches QP at p_0 and Qo' at o'. The line Po' indicates the limiting position beyond which total reflection takes place. If the caustic curve be rotated around the line PQ, the caustic surface will be described.

190. *Refraction through Parallel Layers.*—A ray of light, which is bent out of its original direction on entering a refracting layer (Fig. 101) with parallel plane sides, is, on re-entering the original

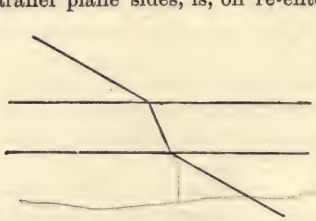


FIG. 101.

medium on the far side of the layer, bent back again into its first direction. This follows at once from the symmetry of the arrangement.

If an object is looked at perpendicularly through such a layer of thickness, t, its distance from the eye is apparently diminished by the amount $t(\mu-1)/\mu$. For, if d be the distance of the object from the layer, an eye situated in the layer would see it as if it were at a distance $d\mu$ from the layer. Therefore, when the eye is just at the far side of the layer, but still inside, the apparent distance of the object is $t+d\mu$. And, if the eye be now just outside the layer, this

distance is decreased in the ratio of μ to unity, which proves the statement.

If a beam of light passes through three such layers the total deviation of the beam from its original direction is the same as if

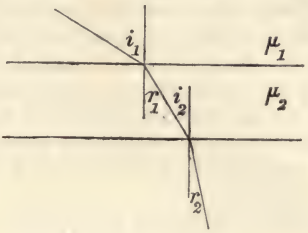


FIG. 102.

the intermediate layer were absent. For, μ_1 and μ_2 being the refractive indices of the first and second media with reference to the second and third respectively we have (see Fig. 102)

$$\frac{\sin i_1}{\sin r_2} = \frac{\sin i_1}{\sin r_1} \cdot \frac{\sin r_1}{\sin r_2} = \frac{\sin i_1}{\sin r_1} \cdot \frac{\sin i_2}{\sin r_2} = \mu_1\mu_2.$$

But, since μ_2 is the index of the second medium expressed in terms of that of the third as the unit, while μ_1 is the index of the first expressed in terms of that of the second as the unit, $\mu_1\mu_2$ is the index of the first medium expressed in terms of that of the third.

It follows that no effect is produced by any number of intermediate layers. For this reason it is possible to calculate the refraction error in the apparent altitude of stars, etc.

191. *Mirage.*—When a beam of light passes non-perpendicularly through a medium composed of parallel layers of continuously

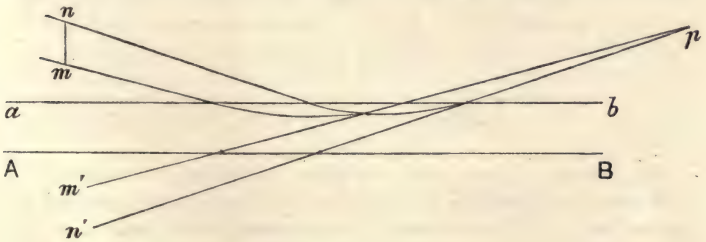


FIG. 103.

varying density, the direction of the beam constantly alters. Let us suppose that, over the surface of the ground AB (Fig. 103) there

exists a stratum of air the density of which continually increases with its distance from the surface. We may suppose also that, above the plane *ab*, the density remains uniform. An object, *nm*, will be seen by an eye situated at *p* by means of rays which enter the non-homogeneous medium and are then bent upwards until, re-entering the uniform medium, they pass straight to *p*. The object appears to be in the direction of *m'n'*, and is obviously inverted since the rays from the upper and lower extremities cross each other before reaching the eye. But the object may also be seen directly through the medium above *ab*, and so the impression of an object with an inverted reflection is produced. This is the case of the ordinary mirage of the desert.

In Fig. 104, AB again represents the surface of the ground, and

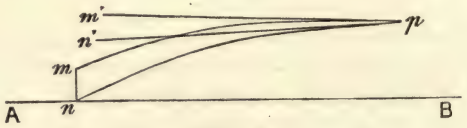


FIG. 104.

the density of the air is suppose to decrease from below upwards. A direct mirage is thus produced.

In Fig. 105 the air is supposed to be of uniform density between AB and *ab*, but is supposed to diminish continuously in density

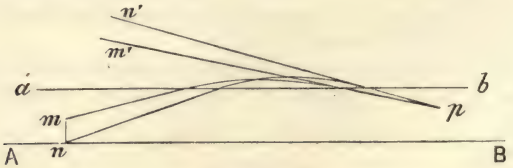


FIG. 105.

above *ab*. This condition obviously gives rise to an inverted image.

192. *Prisms.*—When the two plane bounding surfaces of a medium are not parallel the emergent ray is no longer parallel to the incident ray. Such an arrangement constitutes a prism.

Let ABC (Fig. 106) represent the section, by the plane of the paper, of a prism of some dense medium, such as glass. Let the edge, B, of the prism be perpendicular to the plane of the paper, and let *abcd* be a ray which makes an angle *i* with the (inward) normal

to the face AB and an angle i' with the (outward) normal to the face BC. Let r and r' be the corresponding angles of refraction.

The deviation of the ray from its original direction by the first refraction is $i-r$; and at the second surface this deviation is

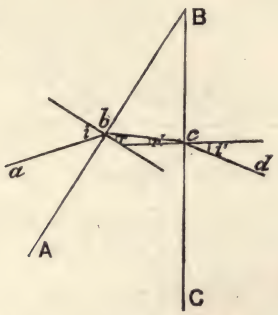


FIG. 106.

farther increased by the amount $i'-r'$. But (§ 188) the alteration of $i-r$, consequent on a given alteration of i, is greater and greater the larger i or r is; and any increase or decrease of r involves an equal decrease or increase of r'. Hence, if i be greater than i', a given diminution of i produces a diminution of $i-r$, which is greater than the simultaneous increase of $i'-r'$. Hence the ray has *minimum deviation* when $i=i'$, that is, when the ray passes through the prism so as to make $Bb=Bc$.

When the angle of the prism is less than $\sin^{-1}1/\mu$ the deviation $i'-r'$ may be in the opposite direction to the deviation $i-r$, but it is easy to see that the above result is true in all cases.

The total deviation, δ, is, in the standard case above, $i'+i-(r'+r)$, which, in the minimum position, becomes $2(i-r)=2i-\alpha$, α being the angle of the prism. If μ be the refractive index of the substance of which the prism is composed, this gives

$$\mu=\frac{\sin(\alpha+\delta)/2}{\sin\alpha/2}.$$

This affords a ready means of determining the refractive index of the substance of the prism; and, if the prism be hollow and have its sides made of glass plates of uniform thickness, the refractive index of any liquid placed in the hollow may be found.

193. *Refraction through Spherical Surfaces: Lenses.* — Let QAB (Fig. 107) represent a part of a spherical surface the centre of

which is at O; and let the refractive index of the substance on the convex side of the surface be μ. Let rays PA, PB, diverging from P, meet the surface in the points A and B respectively. Suppose

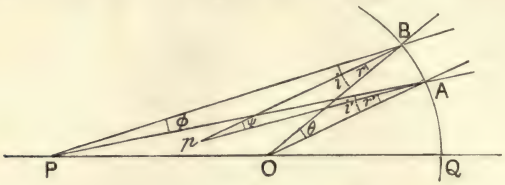


FIG. 107.

that Ap, Bp are the backward prolongations of the refracted rays —μ being assumed to be greater than unity. Join PO and continue the line to meet the surface in Q.

If we denote the angles PAO, PBO, pAO, pBO, BPA, BpA, BOA, by the letters i', i, r', r, ϕ, ψ, θ, respectively, the diagram at once gives $i'-i=\phi-\theta$, and $r'-r=\psi-\theta$. Hence, when the angles are so small that we may write sin $i=i$, cos $i=1$, etc., we have (§ 188) $i'-i=\mu\,(r'-r)$, and therefore

$$(\mu-1)\,\theta=\mu\psi-\phi \quad \ldots\ldots \quad (1).$$

When A coincides with Q and AB is very small this becomes

$$\frac{\mu-1}{OQ}=\frac{\mu}{pQ}-\frac{1}{PQ}.$$

When $\mu=-1$, this gives the first formula of § 183.

Now let the rays (Fig. 108), apparently diverging from the point p, fall upon the second spherical surface of the dense medium at the

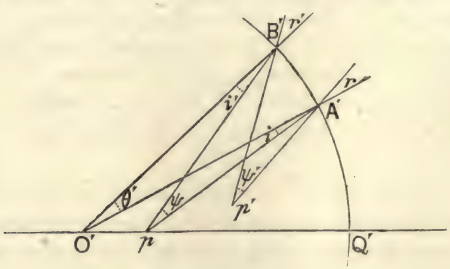


FIG. 108.

points A′, B′. Let O′ be the centre of the new surface, and let O′p meet it at the point Q′. Denote the angles pB′O′, pA′O′, p'B′O′,

$p'A'O'$ $A'p'B'$, $A'O'B'$, by the letters i', i, r', r, ψ', θ', respectively. Since $A'pB'=\psi$, the figure gives $i'-i=\psi-\theta'$ and $r'-r=\psi'-\theta'$. Hence, when $A'B'$ is very small $(\psi'-\theta')=\mu(\psi-\theta')$, or

$$(\mu-1)\,\theta'=\mu\psi-\psi' \ldots\ldots\ldots (2)$$

[Of course (2) might have been deduced from (1) by the substitution of $1/\mu$ for μ.]

From (1) and (2) we get

$$(\mu-1)\,(\theta-\theta')=\psi'-\phi \ldots\ldots (3)$$

When A (Fig. 107) coincides with Q, A′ (Fig. 108) coincides with Q′; and so, when the angles are sufficiently small, we may write (3) in the form

$$(\mu-1)\left(\frac{1}{r}-\frac{1}{s}\right)=\frac{1}{v}-\frac{1}{u} \ldots\ldots (4)$$

r and s being the radii of the two spherical surfaces, while v and u are respectively the distances of p' from Q′, and of P from Q.

The quantity on the left-hand side of equation (4) is constant. We may therefore write (4) in the form

$$\frac{1}{f}=\frac{1}{v}-\frac{1}{u} \ldots\ldots\ldots\ldots (5)$$

where f is a constant, called the *principal focal distance*. When u is infinite, that is, when the incident rays are parallel, we get $f=v$; in other words, f is the distance from Q′ of the point from which originally parallel rays seem to diverge after passing through the medium. This point is called the *principal focus*. P and p' are called *conjugate foci*.

A portion of a medium bounded by two spherical surfaces is termed a *lens*.

In the case just discussed we have taken $s>r$, and have assumed that the concave sides of the spherical surfaces are directed towards P. All lines drawn in the direction PQ are regarded as positive. Lines drawn in the direction of QP are therefore negative.

Equation (4) applies to all lenses. The quantity f is negative (1), when s is positive but less than r, (2) when r is infinite and s is positive, (3) when r is negative and s is positive.

In such cases the principal focus and the source of the parallel rays are on opposite sides of the lens, which is necessarily thickest at the middle (Fig. 109).

But f is positive (1) when r and s are both positive and $s>r$, (2) when s is infinite and r is positive, (3) when s is negative and r is positive.

In the three latter cases the principal focus and the source of the parallel rays are on the same side of the lens, which is now thinnest (Fig. 110) at the middle. Such a lens diminishes the convergence or increases the divergence of the incident rays, while a lens of the

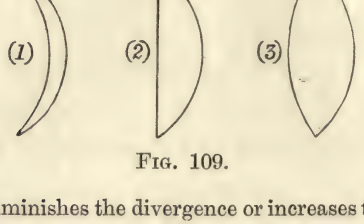


FIG. 109.

previous type diminishes the divergence or increases the convergence of the incident rays. Equation (3) makes this evident at once.

The formula (4) is true of a given small cone of rays which falls perpendicularly upon the lens at its central part, and it shows that,

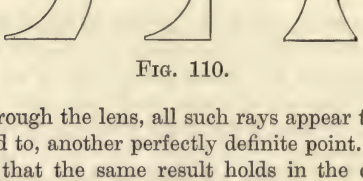


FIG. 110.

after passing through the lens, all such rays appear to diverge from, or are converged to, another perfectly definite point. We shall see in next section that the same result holds in the case of a small obliquely-incident pencil of rays.

194. *Lenses: Oblique Refraction.*—Let DCBA (Fig. 111) represent a ray, which, entering the lens CBQQ′ at C, emerges at B and cuts the line OQ in A. (The letters O, O′, Q, Q′ have the same meanings as in last section.) Let the ray be such that OB and O′C are parallel. Since BC makes equal angles with OB and O′C, it follows that AB and CD make equal angles with them. Therefore AB and CD are parallel.

Let CB be continued to meet OQ in R. We have OR/OB= O′R/O′C, and so the lengths of OR and O′R bear a constant ratio to each other. Therefore R is a fixed point, and is called the *centre of the lens.*

Thus no ray which passes through the centre of the lens is deviated from its original direction by its passage through the lens.

It is easy to find an expression for the distance RQ in terms of the thickness of the lens and its radii of curvature. When the lens is very thin R practically coincides with either Q or Q′.

We shall hereafter assume that we are dealing with thin lenses only.

The final equation of § 189 may be written in the form $op/op' = \mu^2 (1-\sin^2 i)/(\mu^2 - \sin^2 i)$, or, approximately, $\mu^2 (1-i^2)/(\mu^2-i^2)$. This shows that, when i is so small that its square may be neglected, p coincides with p'. And it is easy to prove that, when the reflecting

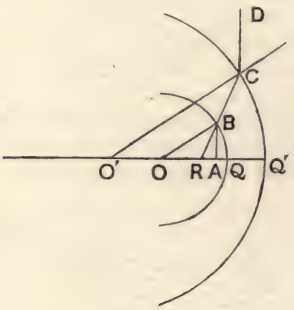


FIG. 111.

surface is spherical instead of plane, p and p' still practically coincide when i^2 is small. (The formulæ of § 193 enable us to calculate the positions of the focal lines.) Hence we conclude that a given, sufficiently small, pencil of rays will, after refraction through a thin lens, be brought to a focus at, or appear to diverge from, one definite point; and this point must be situated on the line drawn from the vertex of the incident pencil through the centre of the lens.

We are now in a position to investigate the production of images by means of lenses.

195. *Formation of Images by Lenses.*—Let AB (Fig. 112) be a thin lens, of which C is the centre, and CF is the principal focal length. If MN be an object which is situated at a distance from the lens greater than CF, rays diverging from N will be brought to a focus at a point n such that $1/CF = 1/CN + 1/Cn$.

Similarly $1/CF = 1/CM + 1/Cm$, and so on.

Thus an inverted real image of MN is formed at mn, and the rays diverging from that image may be examined by an eye which is

situated at the distance of about ten inches from it—that being the usual distance for most distinct vision.

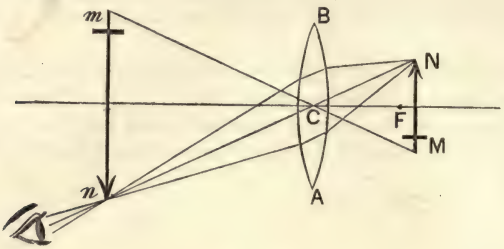


FIG. 112.

On the other hand, if MN be slightly nearer (Fig. 113) to the lens than the principal focus, the lens is only able to diminish the divergence of the incident rays, and the erect and virtual image of MN

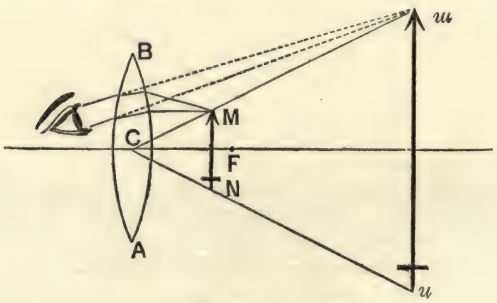


FIG. 113.

is situated at a greater distance from the image than the object is. For the purpose of correct vision as regards an eye placed close behind the lens, this distance must be about ten inches.

The *magnifying power* of the lens is the ratio of *mn* to MN, and is therefore approximately equal to $10/f$, where f is the principal focal length expressed in inches.

The object-glass of the ordinary *astronomical telescope* acts in the manner first described above. The practically parallel rays from a very distant object, MN (Fig. 114) are converged to the principal focus of the object-glass, and so an inverted image, *mn*, is formed. The eye-glass is placed at a distance from *mn* which is slightly less

than its principal focal length, and so forms a magnified image, $m'n'$, still inverted, at a distance of ten inches from the eye.

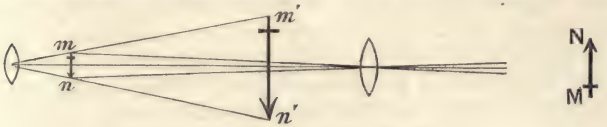


FIG. 114

The magnifying power of the telescope is the ratio of the angle which *mn* subtends at the eye-glass to the angle which it subtends at the object-glass. It is, therefore, approximately equal to the ratio of the focal length of the object-glass to that of the eye-glass.

The arrangement of lenses in the *compound microscope* is essentially the same. The object is placed at a distance from the object-glass which is slightly greater than its principal focal length, and so a magnified inverted image is formed, and is further magnified by the eye-glass.

196. *Dispersion: Aberration.*—In all the preceding sections it was assumed that we were dealing with light of one definite kind or colour alone. But rays of light of different colours are differently refracted by any given substance; and we must, therefore, with this in view, reconsider briefly the action of prisms and lenses.

Let a ray of white light, *ab* (Fig. 115), fall upon a prism, ABC, in a direction perpendicular to its edge. The single ray of white

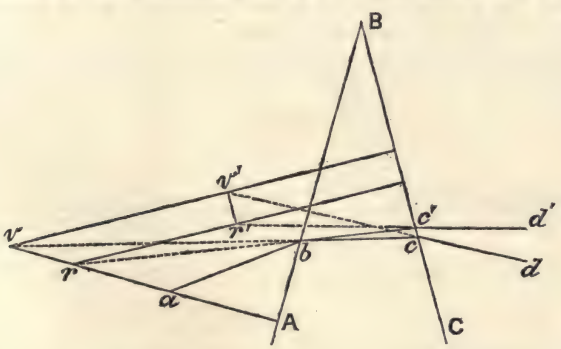


FIG. 115.

light will, on entering the prism, be broken up into a series of coloured rays, *bc*, *bc'*, etc. The red rays are least deviated, and the

blue or violet rays are most deviated, from their original direction. The rays intermediate between the red and the violet are usually broadly characterised in order as orange, yellow, green, blue, and indigo.

Let bc and bc' represent respectively a violet, and a red, ray; and let these lines meet the perpendicular from a on AB in the points r and v. So long as the square of the angle of incidence is negligable, an eye placed in the substance of the prism will see, no white point a but, a coloured line, rv, red at the end nearest to, and violet at the end farthest from, the prism.

After emergence from the face BC, the two rays considered will take the directions cd, $c'd'$. Drop the perpendiculars rr' and vv' on BC, and let dc meet vv' at the point v', while $d'c'$ meets rr' at the point r'. The angles which the emergent rays make with BC being small (which necessitates ABC being small), we see that an eye placed in the same medium as the point a, but on the opposite side of the prism from it, perceives, instead of a, a coloured line, $r'v'$.

Let us suppose now that a represents the section, by the plane of the paper, of a luminous white line, which is parallel to the edge of the prism. In this case $r'v'$ represents a coloured band, which is called a *spectrum*. This separation of the constituents of white light by refraction is termed dispersion. The measure of the dispersion produced by any given substance is the difference of the refractive indices of that substance for the extreme rays of the visible spectrum.

Now let P (Fig. 116) be a luminous white point from which diverging rays fall upon a lens AB. The refractive index of the substance of the lens for violet rays being greater than its refractive index for red rays, the violet rays will be brought to a focus at a

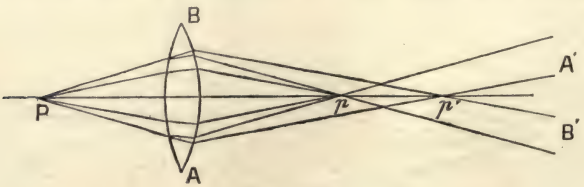


FIG. 116.

point, p, which is nearer the lens than the point p', to which the red rays converge. The light in the regions BpA and B'p'A' is nearly colourless, but, on the whole, it is somewhat violet in the former and reddish in the latter. In the region immediately outside BpA the light is red, while, in the region immediately outside B'p'A', it is violet.

A lens is said to produce *aberration* when it fails to bring all rays diverging from one point to a focus at another point. The aberration is called *chromatic aberration* when it is due to dispersion: it is called *spherical aberration* when it is due to the spherical form of the faces of the lens. The former depends on the first power of i, the angle of incidence, while the latter (§ 194) depends on i^2. So long as the lens is sufficiently thin, and the incidence is sufficiently direct, both are negligable.

197. *Dispersion: Achromatism.*—It is possible to get nearly rid of dispersion while refraction remains, and thus we can obtain a practically achromatic lens. This result can be obtained, since some highly refracting substances produce comparatively small dispersion, while some substances of low refracting power produce comparatively large dispersion—comparatively, that is, to their refraction.

Let us suppose that two prisms, of equal angles, but of different substances, produce dispersions d_1, d_2 respectively, and let their refractive indices for the extreme red light of the spectrum be μ_1, μ_2 respectively. If now we alter the angle of the prism, the dispersion of which is d_1, in the ratio d_2/d_1, and form a compound prism of the two with their edges turned in opposite directions, we shall have a prism which will not produce dispersion of the red and violet rays; but refraction will still take place unless $d_1/\mu_1 = d_2/\mu_2$.

Similarly, we may construct a compound lens for the purpose of avoiding chromatic aberration.

By no pair of substances yet found, however, can we produce complete achromatism. If a lens is completely achromatic for two definite kinds of light it will not be so for any other pair. For if given dispersion is produced by one prism, between a series of pairs of definite kinds of light, equal dispersion will not be produced, by any other prism, between more than one of these pairs. This is known as the *Irrationality of Dispersion.*

Compound lenses, made up of three constituent lenses, can produce closer approximation to achromatism than can a compound lens made up of only two.

198. *Rainbows: Halos.*—We are now in a position to discuss the formation of rainbows and halos.

Let AB (Fig. 117) represent a ray of sunlight which, falling on a drop of water at B, is refracted to C. After reflection at C the ray emerges at D. The whole figure is symmetrical about the line drawn from C through the centre of the drop. The semi-angle between AB and ED is obviously $2r - i$, i and r being respectively the angles of incidence and refraction.

Now (§ 188) this is a maximum when $3 \sin^2 i = 4 - \mu^2$; and, in the case of water and yellow light, $\mu = 1{\cdot}336$. This makes the semi-angle, corresponding to the maximum value of $2r - i$, equal to 21° 1′ nearly. But the existence of a maximum means that the

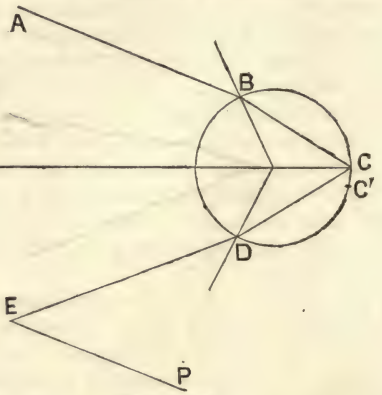


FIG. 117.

rays are crowded closely together in the immediate neighbourhood of the maximum angle, and so an eye situated at E will see comparatively strong yellow light in the direction ED.

Now consider AB to be a ray of white light. This becomes dispersed on refraction, and a blue ray (say) being more refracted, will suffer reflection at a point C′ such that the angle $2r - i$ is smaller than before. Hence the maximum value of the angle $2r - i$ becomes smaller as μ increases.

From E draw EP parallel to AB, and let a very large number of drops be situated symmetrically around EP in the direction of the sun's rays. (Of course, in the figure, the size of the drop is immensely exaggerated relatively to other magnitudes.) The eye at E will see a circle of yellow light, of radius 21° 1′, the centre of which is situated on the line EP. Inside this a circle of blue light will be seen, and, outside it, a red circle will appear—the colours of the spectrum succeeding each other, in order of decreasing refrangibility, from within outwards. This constitutes the explanation, by geometrical optics, of the primary rainbow. In the actual bow the colours are, of course, impure. For the light proceeding from each point of the sun's disc gives rise to a separate bow, and the bow which we see results from the superposition of all these distinct bows.

If the ray AB (Fig. 118) suffers two reflections in the interior of the drop, the emergent ray EP makes an angle with it which is

double of $\pi/2-(3r-i)$. But, § 188, $3r-i$ reaches a maximum when $8\sin^2 i=9-\mu^2$. Hence the angle BPE is a *minimum* when i has the value indicated by this formula. For yellow light, falling on a drop of water, the vertical angle is about 50° 58′.

A ray of higher refrangibility will be reflected at points C′, D′, such that the perpendicular from O on C′D′ will intersect AB at an

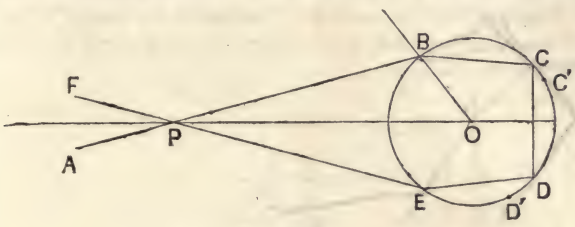


FIG. 118.

angle which is larger than BPO. Hence the minimum value of the angle $\pi/2-(3r-i)$ increases as μ increases. And so, in the secondary rainbow, which is due to two such internal reflections, the colours succeed each other, in order of increasing refrangibility, from within outwards.

Also, since the primary bow corresponds to a *maximum* value of the angle which the emergent ray makes with the incident ray, while the secondary bow corresponds to a *minimum* value of this angle, we see that the space between the bows is devoid of light due to rays which have suffered one or two reflections inside the drops, while there is some such illumination in the region inside the first bow and also in that outside the second.

The bows which are due to more than two internal reflections are too feeble to be visible.

Halos are due to the refraction of light through ice-crystals. The rays are most densely crowded in the directions of minimum deviation. The red rays are least deviated, and therefore appear always on the interior portions of halos. The size of a halo depends upon the effective angle of the ice-crystal to refraction through which it is due. *Parhelia* and *paraselenæ* are simply exceptionally bright portions of halos.

Colourless halos are produced by the reflection of light from the plane surfaces of the crystals.

199. The general phenomena of reflection and refraction receive a ready explanation whether on the corpuscular or on the undulatory theory of light.

Let AB (Fig. 119) represent the bounding surface between two refracting media. Let $p\ q\ r\ s\ t$ represent the path of a corpuscle.

During the rectilinear motion from p to q the corpuscle experiences no resultant attraction in any direction. When it reaches the distance from AB indicated by the line ab, the greater (say) attraction of the medium on the other side of AB preponderates and the path becomes concave towards the surface of separation. This continues until a point, s, at which the resultant attraction again becomes zero, is reached. The rest of the path st is therefore straight, and is inclined at a less angle to the normal to the surface AB than is the part pq.

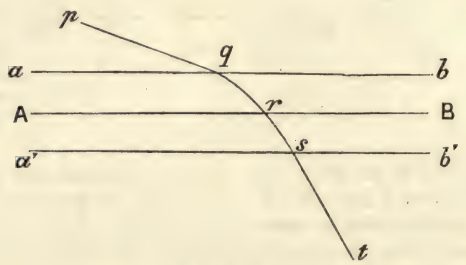


FIG. 119.

Even if the refracting surface is not actually plane, it is yet practically plane, in all cases of finite curvature, so far as the present reasoning is concerned, for the portion qs of the path of the corpuscle is excessively small.

Now (*cf.* § 42) the square of the resolved part of the speed of the corpuscle along the normal increases by a constant amount in passing from ab to $a'b'$, while the resolved part along the surface remains constant, and hence the total speed, v', of the particle in the second medium bears a constant ratio to its total speed, v, in the first. Let, as usual, i and r denote respectively the angles which pq and st make with the normal. We have $v \sin i = v' \sin r$, which is identical with $\sin i = \mu \sin r$, where μ is equal to v'/v. But this is the known law of refraction.

It follows necessarily that the speed of a corpuscle is greater in a dense medium than in a rare one.

200. Let ABC (Fig. 120) represent a plane wave-front, which travelling through the air in the direction indicated by the arrows, reaches the surface, ADF, of a refracting medium. From A, as centre, describe a sphere of radius, AF, such that light will travel over the distance AF in the refracting medium in the same time that it will describe the distance CF in the air. Similarly, from D

describe a sphere of radius, DQ, such that DQ bears to EF (DE is parallel to ABC) the same ratio as the speed of light in the refracting medium bears to its speed in air. It is obvious from the construction that all such spheres touch a plane, PQF, which is therefore the wave-front after refraction.

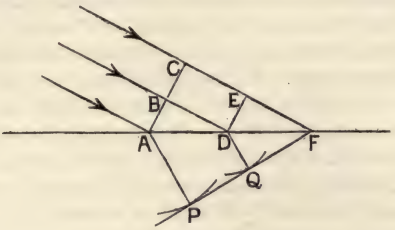


FIG. 120.

Let us denote the speeds of light in the medium, and in air, respectively, by the letters v', v. Let i and r be respectively the angles of incidence and refraction. Then $CAF=i$, $AFP=r$, and $CF=AF \sin i$, $AP=AF \sin r$. But $CF/AP=v/v'$. Hence $\sin i = v/v' . \sin r = \mu \sin r$, if $\mu=v/v'$; and so the known law of refraction is a consequence of the wave-theory of light.

Observe that, in a dense medium in which μ is larger than unity, v' is necessarily less than v. Hence, on this theory, the speed of light must be less in a dense medium than it is in a rare one. This conclusion is in direct opposition to that derived from the principles of the corpuscular theory; and so we are furnished with a *crucial test* between the two theories. The result of experiment is (§ 178) entirely in favour of the undulatory theory. Consequently, we shall hereafter deal with the results of this theory alone.

It is easy to deduce from the above theory the fact that the time taken by light to travel from a given point in one medium to a given

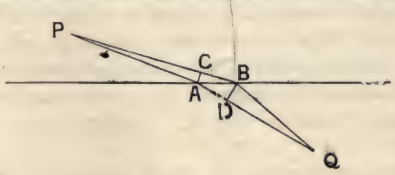


FIG. 121.

point in another is a minimum. Let PAQ (Fig. 121) be the actual path of light proceeding from P to Q, and let PBQ be a *very near* path. Draw AC and BD perpendicular to PB and AQ respectively,

Then $CB/AD = \sin i/\sin r = v/v'$. That is, $CB/v = AD/v'$, or AD and CB are described in equal times. Therefore, and since $PA = PC$, $DQ = BQ$, the paths PAQ and PBQ are described in equal times. But, as B moves away uniformly from A, $QA - QB$ increases at a diminishing rate, while $PB - PA$ increases at an increasing rate. Hence the time of description of PAQ is a minimum.

This law of least time was first given by Fermat. It is true in the case of reflection also (see § 182). The corpuscular theory gives $PAv + AQv' =$ a minimum. This sum is termed the *action*.

201. *Hamilton's Characteristic Function.*—In the earlier part of the present century Sir W. R. Hamilton introduced a general method by which all optical problems may be solved by a single process. The following extracts from an 'Account of a Theory of Systems of Rays,' written by Hamilton himself, and published in his *Life* will indicate the nature of his work.

'A *Ray* in optics is to be considered here as a straight or bent or curved line, along which light is propagated, and a *System of Rays* as a collection or aggregate of such lines, connected by some common bond, some similarity of origin or production, in short, some optical unity. Thus the rays which diverge from a luminous point compose one optical system, and, after they have been reflected at a mirror, they compose another. To investigate the geometrical relations of the rays of a system of which we know (as in these simple cases) the optical origin and history, to inquire how they are disposed among themselves, how they diverge or converge, or are parallel, what surfaces or curves they touch or cut, and at what angles of section, how they can be combined in partial pencils, and how each ray in particular can be determined and distinguished from every other, is to study that System of Rays. And to generalize this study of one system so as to become able to pass, without change of plan, to the study of other systems, to assign general rules and a general method whereby these separate optical arrangements may be connected and harmonized together is to form a *Theory of Systems of Rays*. Finally, to do this in such a manner as to make available the powers of the modern mathesis, replacing figures by functions and diagrams by formulæ, is to construct an Algebraical Theory of such Systems, or an *Application of Algebra to Optics*. . . . .

'The method employed in that treatise (Malus's *Traité d'Optique*) may be thus described: The direction of a straight ray of any final optical system being considered as dependent on the position of some assigned point upon that ray, according to some law which characterizes the particular system and distinguishes it

from others; this law may be algebraically expressed by assigning three expressions for the three co-ordinates of some other point of the ray, as *functions* of the three co-ordinates of the point proposed. Malus accordingly introduces general symbols denoting three such functions (or, at least, three functions equivalent to these), and proceeds to draw several important general conclusions, by very complicated but yet symmetrical calculations, many of which conclusions, along with many others, were also obtained afterwards by myself, when, by a method nearly similar, without knowing what Malus had done, I began my own attempts to apply algebra to optics. But my researches soon conducted me to substitute, for this method of Malus, a very different, and (as I conceive I have proved) a much more appropriate one, for the study of optical systems; by which, instead of employing the *three* functions above mentioned, or at least their *two* ratios, it becomes sufficient to employ *one function*, which I call *characteristic* or principal. And thus, whereas he made his deductions by setting out with the *two equations of a ray*, I, on the other hand, establish and employ the *one equation of a system*.

'The function which I have introduced for this purpose, and made the basis of my method of *deduction* in mathematical optics, had, in another connection, presented itself to former writers as expressing the result of a very high and extensive *induction* in that science. This known result is usually called the *law of least action*, but sometimes also the principle of *least time*, and includes all that has hitherto been discovered respecting the rules which determine the forms and positions of the lines along which light is propagated, and the changes of direction of those lines produced by reflection or refraction, ordinary or extraordinary. A certain quantity which in one physical theory is the *action*, and in another the *time*, expended by light in going from any first to any second point, is found to be less than if the light had gone in any other than its actual path, or at least to have what is technically called its variation null, the extremities of the path being unvaried. The mathematical novelty of my method consists in considering this quantity as a *function* of the co-ordinates of these extremities, which varies when they vary, according to a law which I have called the *law of varying action;* and in *reducing all researches respecting optical systems of rays to the study of this single function:* a reduction which presents mathematical optics under an entirely novel view, and one analogous (as it appears to me) to the aspect under which Descartes presented the application of algebra to geometry.'

CHAPTER XVII.

RADIATION AND ABSORPTION: SPECTRUM ANALYSIS. ANOMALOUS DISPERSION. FLUORESCENCE.

202. We have already discussed the reflection (including scattering) and refraction of light at the common surface of two media. We have now to consider the absorption of light in its passage through material media, together with other associated phenomena.

According to the undulatory theory (which we now assume to be true, and of the truth of which we shall receive additional evidence as we proceed) light consists of waves propagated through a medium (called the *ether*) which fills space.

The particles of a body which is emitting radiation must therefore be in rapid vibratory motion, and must communicate their motion to the ether. The parts of the body, the vibrations of which are communicated to the ether, may be the molecules, or the constituent parts of the molecules, or even the atoms.

When a bell is struck violently and frequently the resulting sound is extremely complex and consists of notes, of various pitches, which may differ greatly from each other in intensity. The more violent the blows upon the bell become, and the more rapidly they are made, the more complex will the clang be. New forced vibrations appear, and the intensity of each of those previously existing is increased. It is only when the blows are excessively feeble and unfrequent that the fundamental tone is heard by itself.

Now the molecules of a solid body, of high temperature, are constantly colliding; and the vibrations induced in a molecule by one collision do not die out before another is sustained. Hence the radiation given off by such a body consists of vibrations of many periods; and, as the temperature of the body becomes higher, vibrations of shorter and shorter period make their appearance, and the intensity of all previously existing vibrations becomes greater. Thus, if we examine the spectrum of a body the temperature of which is gradually raised, we may at first perceive no luminous

radiation at all; but, as the temperature rises, first red light will appear, then yellow, and so on, until a complete continuous spectrum is seen, and the luminosity of every part gradually increases as the temperature rises.

In the case of an ordinary gas, however, the molecules usually are sufficiently free from collisions to allow of them vibrating in their own proper modes. Hence radiations of definite periods only will be emitted; and thus the spectrum of a gas is discontinuous, and consists of bright lines. It varies with the temperature and pressure. As the pressure is increased the lines broaden out, and the spectrum gradually becomes continuous, like that of a liquid body or of a solid.

We already know (§ 173) that a body which has a definite period of vibration, and which is at rest, may be set in vibration by the communication to it, through an intervening medium, of vibrations, of its own proper period, which are emitted by another body. Hence, if radiation travelling through the ether enters a material medium (solid, liquid, or gas), and if the natural period of oscillation of the molecules coincides with the period of some of the ethereal vibrations, the molecules will be set in motion, and the energy of the radiation will be diminished.

This is the process which is termed *absorption*.

In many cases (if not really in most cases) the period of the induced vibration is longer than that of the ethereal vibration which induces it (§ 208). In almost all cases the absorbed energy is manifested by a rise of temperature.

203. *Equality of Emissivity and Absorptive Power.* The *Absorptive Power* of a body, under given conditions, for any definite radiation, is the fraction of the whole incident radiation of that kind which it absorbs. Now a black body is one which absorbs all the incident radiation; so we might define the absorptive power for the given radiation as the ratio of the amount of it which the body absorbs to the amount of it which a black body would absorb.

The *Emissivity* of a body, at a given temperature, for any given radiation, is the ratio of the quantity of that radiation which it emits to the quantity of it which is emitted by a black body under the same conditions.

An extremely simple relation connects these quantities: *The Emissivity and Absorptive Power of a body, at a given temperature, for any radiation, are equal.*

The proof of this law was given by Stewart in 1858. The law is (see § 255) an extension of the statement, made by Prevost about a century ago, that the radiation emitted by a body depends solely

upon the nature of the body, and upon its temperature; and various experimental illustrations of it were known long before Stewart's proof was given. Brewster had shown that definite portions of the sun's light are absorbed in its passage through the earth's atmosphere. Foucault had pointed out that, while the electric arc emits (more freely than its other radiations) yellow light of two definite refrangibilities, the light from one carbon pole is robbed of these two kinds of radiation when it passes through the arc. Stokes also had explained this by the analogous properties of sounding bodies (§ 173).

It is known, as an experimental result, that a number of bodies, at different temperatures, placed inside an enclosure which neither allows radiation to pass outwards from within it nor inwards from without it, will ultimately arrive at, *and maintain*, one common temperature. But this could not result unless each body emitted radiation at precisely the same rate as that at which it absorbed the radiation. This proves the law so far as radiation *as a whole* is concerned. [No such enclosure as has been postulated exists in nature; but a polished reflecting surface of silver would form a sufficiently close experimental approximation; and, the more nearly the condition is satisfied, the more nearly does the result hold.]

The radiation inside the enclosure must be that of a black body at the same temperature, for any one of the bodies might be a black one. Let us suppose that one of the bodies absorbs one definite radiation only, and allows all others to pass freely through it. (Solutions of didymium salts approximately possess this property.) This body must emit the same kind of radiation as it absorbs, and that to precisely the same extent; otherwise its temperature would vary. This proves the law as stated for any definite radiation.

Many experimental illustrations of the truth of the law were given by Stewart. Thus, a piece of red glass, held in front of a fire, appears red because it absorbs the green and blue rays. If placed in the fire it becomes colourless when its temperature becomes equal to that of the fire—for it then still allows the red rays to pass through it, and, in addition, itself emits the rays which it absorbed. If taken out of the fire and held in a dark room it emits bluish-green light—precisely that which it absorbed.

Again, Stewart, and also Kirchoff (who arrived at the results under consideration independently of, though somewhat later than, Stewart), showed that a plate of tourmaline, cut parallel to the axis of the crystal, emits, when heated, light which is polarised (Chap. XIX.) perpendicularly to that which it allows to pass, that is, it emits the rays which it absorbs when cold.

204. *Spectrum Analysis.* — In order to examine the luminous radiation emitted by a given body, we may place in front of the body a narrow vertical slit, which is situated at the principal focus of a convex lens. The light diverging from the slit is thus condensed into a parallel beam which is passed through a prism (usually of dense glass) and so gives rise to a spectrum. This spectrum is magnified by means of a telescope. Such an arrangement essentially constitutes the instrument called a *spectroscope* (or *spectrometer*, if a graduated circle and vernier are attached for the purpose of determining the angular positions of the telescope when it is directed towards different parts of the spectrum).

Let us suppose that we are examining the light emitted from a highly-heated lime-ball, and that this light, before falling on the slit, passes through a Bunsen flame in which metallic sodium is being vaporised. The lime-ball alone would give a continuous spectrum. The sodium-tinged flame alone would give a discontinuous spectrum consisting of two bright yellow or orange lines situated close together. The spectrum actually observed is continuous, but has two bright lines in the same position as those in the spectrum of the sodium flame.

If we vaporise the sodium in the flame of a spirit lamp instead of in a Bunsen flame, everything else remaining the same, a continuous spectrum, crossed by two *dark* lines in the positions of the former bright ones, will be seen. This was pointed out and explained by Kirchoff.

The cause lies in the difference of the temperatures of the Bunsen flame and the flame of the spirit lamp. A line will appear bright, or dark, according as the intensity of the radiation of that particular kind from a black body at the temperature of the flame exceeds, or falls short of, the intensity of the radiation of that kind which is emitted by the source. For, if R be the intensity of the given radiation as emitted from the source, while $R' = pR$ is the intensity of the light of that kind emitted from a black body at the temperature of the flame, and ρ is the radiating power (or absorptive power) of the flame for that radiation, the intensity of the given kind of light which reaches the eye is $R - \rho R + \rho p R = R[1 + \rho(p-1)]$. This quantity exceeds, or falls short of, R, according as p is greater, or less than, unity. If the source were a black body, p could not exceed unity unless the temperature of the flame were greater than that of the source; but, the source not being a black body, p may exceed unity, although the temperature of the flame is below that of the source, *i.e.*, bright lines may be visible. If, however, the difference of temperature of the source and flame be sufficiently great, the lines will appear dark.

The spectrum of sunlight was found by Wollaston and (later) Fraunhofer to exhibit a number of persistent dark bands. In accordance with the above principles we conclude that these lines are due to absorption. Some of them can be shown to be due to absorption in the earth's atmosphere, but the great majority are produced by absorption in the (comparatively) cold vapours surrounding the hot body of the sun. Sufficient matter to produce such absorption does not exist in the space between the earth and the sun.

Now we can experimentally determine the kinds of radiation emitted by the hot vapours of the various elementary substances. And if it is found that any of these radiations are absent from the spectrum of sunlight it is to be inferred that the vapours of these substances (provided the cause is not terrestrial) are present in the regions immediately surrounding the sun. In this way it is found that a great many substances existing on the earth's surface are

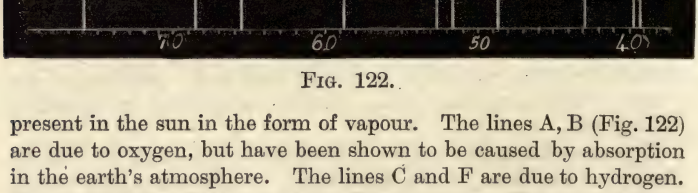


FIG. 122.

present in the sun in the form of vapour. The lines A, B (Fig. 122) are due to oxygen, but have been shown to be caused by absorption in the earth's atmosphere. The lines C and F are due to hydrogen. The (double) line D is caused by sodium vapour. The (triple) line *b* is produced by the vapour of magnesium. Some hundreds of lines are caused by the presence of the vapour of iron.

In the same way the light emanating from any star, comet, or nebula, etc., may be examined, and the chemical constitution of the luminous body inferred. The various stars may be classified into several groups according to the nature of their spectra, and this classification indicates approximately their relative age. Generally speaking, the more recent stars have bright-line spectra, while the older stars exhibit continuous spectra crossed by numerous dark lines. The spectra of nearly extinct stars, however, resemble those of recent stars to a considerable extent.

If the slit of the spectrometer is wide the various coloured images overlap and produce an impure spectrum ; and, under the same condition, a bright light broadens out and becomes indistinct. But, however narrow the slit may be, a line—even if due to light of one definite refrangibility alone—has always some finite breadth. The

reason is that the radiating molecules are in violent motion—some moving towards, others moving from, the observer.

If n be the number of vibrations produced in the ether per second, and, if the molecule emitting the light be at rest relatively to the observer, the wave-length, λ, of the disturbance is given by the equation $V = n\lambda$, where V is the speed of light. But, if the molecule be moving, relatively to the observer, with speed $\pm v$, the wave-length will be given by the equation $V \pm v = n\lambda'$; and the apparent change of wave-length is

$$\lambda' - \lambda = \pm v/n.$$

Hence, the molecules of a luminous body having all possible speeds included between the limits $+v$ and $-v$, a bright line in the spectrum of its light will possess finite breadth, even when it corresponds to one definite kind of radiation alone.

This principle has been applied to determine the rate of rotation of the sun on its axis, the speed of projection of gases in a solar eruption, and the rate of motion of stars to or from the earth.

205. *Law of Absorption: Body Colour. Dichroism.*—Let R be the amount of radiation of some definite kind which falls upon an absorbing medium. Let ρ (called the *absorption co-efficient*) be the percentage of this radiation, which is stopped by a plate of the medium of unit thickness. The quantity which passes through the given plate is therefore $R(1-\rho)$. A second plate of the substance, also of unit thickness, will stop the fraction, ρ of this quantity; so that the amount, $R(1-\rho)^2$, passes through a plate the thickness of which is two units. And, generally, the quantity which passes through a plate, the thickness of which is n units, is $R(1-\rho)^n$. This practically vanishes, however small ρ may be (provided only that it is finite), when n is sufficiently great. Conversely, the amount of radiation, of the given kind, from a sufficient thickness of such a substance, is equal to that of a black body at the same temperature. [The expression $R(1-\rho)^n$ is only true on the assumption that ρ is constant for all radiations considered. If it is not so, we must write $\Sigma \cdot R(1-\rho)^n$ instead.]

A substance which absorbs (say) red light, will appear bluish-green when viewed by transmitted light. And, the greater the thickness of the substance through which the light passes, the denser will be the apparent colour, until, finally, practically no light can pass.

Hence a substance into which light penetrates for a short distance and is then reflected out, will appear to be coloured, provided that selective absorption takes place, and its colour will be the same as

that of the light which it transmits. This colour is termed the *body colour* of the substance.

For example, a mixture of blue and yellow pigments appears to be green because, if white light falls upon it, the particles of the blue pigment absorb the rays of small refrangibility, while the particles of the yellow pigment absorb the rays of large refrangibility. The green rays alone are partially reflected by both substances, and so the mixture appears to be green. (A mixture of blue and yellow *lights* is of a purplish colour. This may be seen by rotating rapidly a disc, divided into sectors, some of which are coloured yellow, and some blue.)

Now suppose that some substance absorbs (say) red light and green light, and let the coefficient of absorption for red light be much greater than the coefficient of absorption for green light. Suppose also that, in the incident light, the red rays are more intense than the green rays. It is obvious that, while the intensity of the red rays in the transmitted light will exceed the intensity of the green rays so long as the thickness of the substance is small, after a certain thickness is reached, the green light will be transmitted in greater intensity than the red light. The colour of such a substance will therefore change from a reddish hue to a greenish hue, as seen by transmitted light, as its thickness increases. This phenomenon is known as *dichroism*.

The accompanying diagram illustrates these facts graphically. The abscissæ of the curves represent the thicknesses of the absorbing medium; the ordinates of one set of curves represent the intensities of the transmitted light of one kind, and the ordinates of the other set indicate the intensities of the transmitted light of another kind, corresponding to the various thicknesses. The numbers accompanying the curves indicate different values of the coefficients of absorption. At the point p the high absorptive power (0·7) of the substance for the originally more intense light has diminished the intensity of that light to the same value as that which is exhibited by the originally feeble light, for which the co-efficient of absorption is only 0·1.

Many such bodies occur in nature. For example, glass, coloured with a cobalt salt, while it transmits blue light when its thickness is small, appears red by transmitted light when its thickness is sufficiently great.

The law of absorption, above stated, *must* be true (neglecting such extraneous effects as internal reflection or scattering of light) so long as the coefficient of absorption does not depend upon the intensity of the light. Such experiments as have been made to test this point furnish confirmatory evidence.

206. *Surface Colour: Metallic Reflection.*—Some substances reflect from their surface certain rays only; thus gold reflects yellowish rays, and copper reflects reddish rays. The colour produced by this '*metallic reflection*' is called *surface colour.*

The light which is transmitted by a thin film of such substances is complementary to that which is reflected, that is, the transmitted light and the reflected light together make up a light of the same

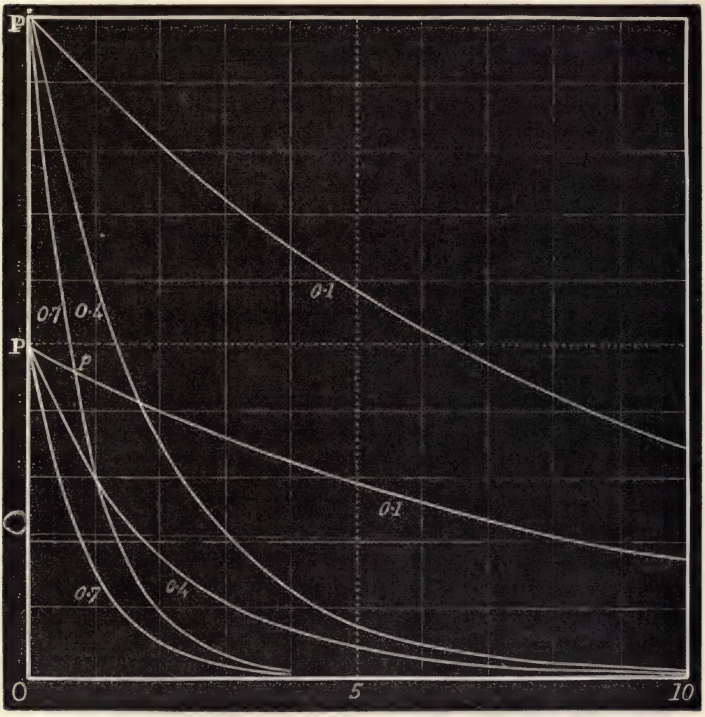


FIG. 123.

composition as that which was incident upon the surface. The reflected light cannot be plane-polarised at any angle of incidence (Chap. XIX.).

Many substances, besides metals, exhibit surface colour—for example, thin films of rose aniline, or of blue aniline, etc., appear of different colours according as they are viewed by transmitted, or by reflected, light. Such films may be prepared by placing a layer of

an alcoholic solution of the aniline on a plate of glass and allowing the alcohol to evaporate. The colour, as seen by reflected light, varies somewhat with the angle of incidence.

The light reflected from such films cannot be entirely polarised at any angle of incidence. It consists of two parts—a part which can be plane-polarised at a certain angle of incidence and is identical with the transmitted light (which, in fact, constitutes the body colour of the substance)—and a part which cannot be plane-polarised, and so resembles the surface colour of metals. The polarisable part may be got rid of by suitable means, so that the remaining part may be examined alone. In the case of permanganate of potash, Stokes found that the surface colour seemed to be due to precisely those rays which were absent from the transmitted light, or, which is the same thing, the body colour. Hence, the colour of the light transmitted through this substance is due only to a very slight extent, if at all, to absorption. The spectrum of the transmitted light has five dark bands in the green part; the reflected light is green, and the spectrum of the surface-colour portion of it consists of five bright bands, which correspond to the dark bands in the spectrum of the transmitted light.

207. *Anomalous Dispersion.*—In close association with the existence of dark absorption bands appears the phenomenon of *anomalous*, or *abnormal* dispersion.

In general, the rays of greater wave-length suffer refraction, on passage through a prism, to a smaller extent than the rays of shorter wave-length. But, in many substances, this rule does not hold. Such media are said to possess the property of anomalous dispersion.

Fox Talbot was the first to observe the phenomenon, but he did not publish his observations for about thirty years. In the meantime Le Roux has observed that iodine vapour refracted red light more than it refracted blue light.

Christiansen, Kundt, and others have widely extended our knowledge of such substances. Kundt has shown that the property of anomalous dispersion is possessed by all substances which exhibit surface-colour.

If a continuous spectrum, such as may be given by a glass prism, be examined through another prism of a substance which exhibits abnormal dispersion, the spectrum will no longer be continuous but will present one or more dark bands. If the second prism be now turned so as to have its edge at right angles to the edge of the glass prism, the parts of the continuous spectrum will be displaced from their original positions to an extent depending upon the refractive

index of the substance for each kind of light. The displacement of the rays, in a part of the spectrum close to a dark band, but of smaller wave-length than the absorbed rays, is abnormally small; and the displacement of the rays of slightly larger wave-length than those which are absorbed is abnormally great (Fig. 124).

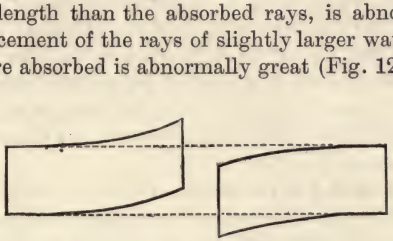


Fig. 124.

The general law, as given by Kundt, is that the rays of slightly less refrangibility than the absorbed rays have their refrangibility abnormally increased, while the rays of slightly greater refrangibility than the absorbed rays have their refrangibility abnormally diminished on passage through the absorbing medium.

208. *Fluorescence.* — The phenomenon of fluorescence is also necessarily associated with the absorption of light.

Brewster observed that the path of a beam of white light through a solution of chlorophyll glows with red light, and he termed the phenomenon 'internal dispersion.' Then Herschel noticed that the surface of a solution of sulphate of quinine upon which sunlight falls is of a bright blue colour. He named this appearance 'epipolic dispersion'; but Brewster showed that the blue colour could be manifested in the interior of the liquid if the beam of sunlight were sufficiently concentrated, and so he concluded that the phenomenon was of the same kind as that which he had already observed in the case of chlorophyll and fluorspar, etc.

Stokes has shown that a great many ordinary substances, such as bone, white paper, etc., possess this property, to which (avoiding any reference to dispersion) he gave the name of *fluorescence*, from its being noticeable in fluorspar. His method of observation consisted in allowing a beam of light to enter a darkened chamber through a plate of blue cobalt-glass. This beam fell partly upon a white non-fluorescent body (white porcelain), and partly upon the body under examination. The light reflected from the two substances was then examined through a slit and prism. In this way the fluorescent light was compared with the light which produced it.

Stokes found that the light which was emitted by the fluorescent body was always of lower refrangibility than the light which pro-

duced it. The fluorescence of sulphate of quinine is due to the extreme violet rays of the spectrum, and to invisible rays of still higher refrangibility. Hence, by means of such a solution, the absence of rays beyond the visible part of a spectrum may be determined. For, if the spectrum be thrown on a screen damped with this solution, fluorescence is produced, beyond the usual visible part, except when rays of certain refrangibilities may be absent. In the case of chlorophyll the light which produces the effect is chiefly in the visible spectrum.

The explanation of the phenomenon given by Stokes is that the ethereal vibrations are absorbed by the fluorescent matter, which is set in vibration, the period of its vibration being usually longer than, never shorter than, the period of vibration of the ether. The vibrating matter now reacts upon the ether, and sets up in it vibrations which are generally longer still, but are never shorter than those induced in the molecules of the matter. This explains the lowering of refrangibility.

Dynamical illustrations of such interaction can be given. The following is due to Stokes. Ships at rest on a calm sea may be set in vibration by waves of definite period propagated from a distance. The natural period of oscillation of each ship will not generally agree with that of the waves. Any ship which is thus set in vibration will, by its vibrations, produce waves which spread outwards from it; but the period of these waves will generally be greater than that of the original waves, and can never be less than it.

If Stokes's explanation be true it is to be expected that the light which gives rise to fluorescence will be absent from the absorption spectrum of the substance. This is invariably the case.

Phosphorescence is precisely the same phenomenon as fluorescence. The only difference which subsists between the two is a difference of duration. Phosphorescence (so-called) frequently lasts for hours after the stimulating radiation is removed; fluorescence is maintained usually for only a small fraction of a second after the light ceases to fall on the substance.

Becquerel demonstrated and measured the finite time of duration of fluorescence in many substances. His apparatus consisted of a box with perforated revolving discs at either end. The perforations were so arranged that one end of the box was closed, while the other was open. The substance which was to be tested was placed inside the box, and, on the discs (which had a common axis) being rotated, an intermittent beam of light passed through the substance. No light could pass out at the end of the box opposite to that at which the light entered unless the substance were fluorescent.

But, that condition being satisfied, light could pass through when the speed of rotation of the discs was sufficiently great.

The duration of fluorescence is exemplified in the above dynamical illustration by the continued oscillation of the ships for some time after the cessation of the disturbance which originates it.

209. *Theories of Dispersion.*—Cauchy was the first to advance a dynamical theory of dispersion. He ascribed it to the coarse-grainedness of the matter of which the dispersing substance is composed. The great difficulty of this theory is that, in order to account for the observed values of the refractive indices of substances such as glass, etc., the number of molecules of matter existing side by side in the length of a wave of light must be assumed to be much smaller than, from other considerations (§ 146), can possibly be admitted. Sir W. Thomson has recently shown that Cauchy's hypothesis can be so modified as to enable it to surmount this difficulty.

This hypothesis leads to an expression for the refractive index, μ, of any substance of the form

$$\mu = a + \frac{b}{\lambda^2} + \frac{c}{\lambda^4} + \ldots\ldots$$

where a, b, and c, etc., are constants, and λ is the wave-length. This formula shows that the refractive index increases as the wave-length diminishes. Its results accord very well with experimental observations within the range of the visible spectrum, but it does not apply well to the invisible rays at the less refrangible end of the spectrum. The various terms rapidly diminish in numerical magnitude.

Briot generalised Cauchy's investigation somewhat, and deduced the expression

$$\frac{1}{\mu^2} = x\lambda^2 + a + \frac{b}{\lambda^2} + \frac{c}{\lambda^4} + \ldots\ldots$$

which agrees better with experimental observations than the former does, and applies to a much greater range of wave-lengths.

The term $x\lambda^2$ depends upon the direct action assumed to exist between the ether and matter.

Modern theories (for example, that of v. Helmholtz) have regard, not so much to space relations—between wave-length and molecular distance—as to time relations—between the period of vibrations in the ether and the period of free oscillation of the material molecules.

V. Helmholtz assumes the existence of a viscous resistance to the motion of the molecules. When the periods of the ethereal and the molecular vibrations are identical, or approximately identical,

absorption takes place, and, because of the viscosity, the vibrational energy takes the form of heat.

Thomson's results differ from those of v. Helmholtz chiefly because he purposely avoids the assumption of the existence of viscosity. He obtains the equation

$$\mu^2 = 1 + \frac{c_1\tau^2}{\rho}\left(-1 + \frac{q_1\tau^2}{\tau^2 - x_1^2} - \frac{q_2\tau^2}{x_2^2 - \tau^2} - \ldots\right)$$

when μ is the refractive index, τ is the period of vibration of the ether, and x_1, x_2, etc., are the natural periods of oscillation of the molecules arranged in ascending order of magnitude. So long as τ is considerably greater than x_1 and considerably less than x_2, this equation will correspond to the case of ordinary refraction. As τ approaches x_1 in value the refractive index is abnormally increased. When τ is less than x_1, μ^2 is at first negative, but afterwards becomes positive, though abnormally small, as τ still further decreases. This explains the existence of anomalous dispersion. Negative values of μ^2, which accompany anomalous dispersion, indicate the existence of absorption or metallic reflection.

Thus the high reflecting power of silver is, on this theory, due to the fact that each one of all the kinds of radiation which are observed to be reflected from it has a vibrational period which is smaller than the smallest of the natural periods of oscillation of the molecules of silver.

Again, when τ has such a value that μ^2 is positive, but is less than unity, the particular radiation, of which τ is the period, will pass through the substance more quickly than it passes through air.

The energy of the rapid vibrations of the molecules is gradually transmuted into energy of the slow vibrations. This explains fluorescence and the radiation of heat from a body which has absorbed light. The molecule may be so constituted that the fluorescence (or phosphorescence) may last for a very long time.

CHAPTER XVIII.

INTERFERENCE. DIFFRACTION.

210. *Principle of Interference.*—If light consists of undulations, propagated through the ether, the effects of which, at any point of the ether, are superposed in precisely the same way as are the effects of separate simple harmonic motions (§ 52), we should expect that conditions might occur under which the resultant motion at that point would be null—while, under other conditions, the resultant motion might be exceptionally great. We already know that, for a similar reason, when waves are propagated along the surface of water from two different sources, no resultant disturbance of the surface may exist at certain points. So also sounds from two different sources may be totally unheard by an ear placed at certain positions within hearing distance.

In order to produce continuous interference at given points it is absolutely necessary that the waves diverging from two sources should be of precisely the same period, as otherwise the resultant disturbance would vary from a minimum to a maximum alternately. Thus, in the case of sound, difference of period gives rise to beats which may be observed by the ear.

Now the phase of the vibration emitted from one point of a flame has absolutely no relation with the phase of that emitted from any other point; and hence we cannot expect observable interference between rays coming from different luminous sources. Interference of course does occur between such rays constantly, but, in general, the alternations between maximum and minimum effects will succeed each other so rapidly that the eye can perceive no variation of intensity.

Therefore we conclude that, in order that persistent interference effects may be observable, the two interfering rays must originally have proceeded from a common source.

More than two centuries ago Grimaldi observed that, when rays of light from two sources overlapped each other and fell

upon a screen, the portion of the screen which was illuminated by the two rays appeared to be darker than when it was illuminated by one ray alone. He allowed sunlight to enter a darkened chamber through two small apertures in the shutter. But these apertures were illuminated by light coming from all portions of the sun's disc, and so the effect which Grimaldi observed, to whatever cause it may have been due, could not have been produced by interference. Grimaldi, indeed, was not looking for interference phenomena—this was not thought of until 150 years later—he wished to prove that light was not material, since two portions of light apparently destroyed each other. And this reasoning is practically conclusive, for the conditions which would have to be assumed, in order to make an explanation of these phenomena by the emission theory possible, would be so arbitrary and artificial that no one could seriously advance them.

211. *Young's Experiment.*—Young was the first to observe true interference effects. He admitted light through a single small aperture in a shutter behind which he placed another shutter pierced by two small openings. In this way he obtained two rays of light which proceeded originally from a common source—the single opening in the first shutter. That portion of a screen which was illuminated by both rays was crossed by alternately-arranged dark and bright bands. Young observed that the bands became narrower when the distance between the holes in the second screen was increased. He also noticed that the effect disappeared if either opening were closed.

The wave theory affords a ready explanation of the phenomena which Young observed.

Let A, A′ (Fig. 125) represent the two openings in the screen, and let AP, A′P be two rays which each illuminate the point P of the

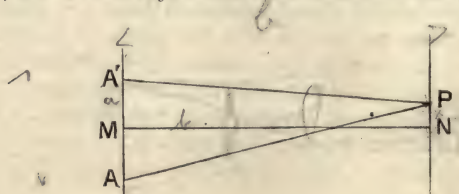


FIG. 125.

screen PN. M is the central point of AA′, and MN is drawn perpendicular to AA′ and PN. Denote the length of AM (or A′M) by a and the length of MN by b, and let x represent the distance PN.

The waves, which travel along AP and A′P, start from A and A′ respectively in the same phase. Consequently the point P will be bright or dark according as AP − A′P is an even, or an odd, multiple of half a wave-length.

Now $AP^2=(a+x)^2+b^2$ and $A'P^2=(a-x)^2+b^2$. Therefore $AP^2-A'P^2=(AP+A'P)(AP-A'P)=4ax$. But ($a$ and x being very small in comparison with b) AP + A′P is approximately equal to $2b$, and so the condition gives

$$\frac{2ax}{b}=n\frac{\lambda}{2},$$

and the point P is bright or dark according as n is an even or an odd integer.

This formula indicates that, to the degree of approximation with which we are dealing, the locus of P, when b varies and n is constant, is the straight line MP. The *exact* locus is a hyperbola, of which A and A′ are the foci. This follows at once from the condition AP − A′P = a constant.

A and A′ may, of course, represent narrow luminous strips with their length perpendicular to the plane of the paper. The point P then corresponds to a dark or bright band also perpendicular to the plane of the paper.

By measuring the quantities a, b, and x, and by counting the number, n, of the particular band under observation, we can calculate the value of λ.

The distance between the n^{th} and the $(n+1)^{th}$ band is independent of n, and is therefore constant when a, b, and λ are fixed.

212. *Fresnel's Experiment.*—In Young's experiment the beams of light passed through apertures cut in a solid. Hence the observed effects might have been due to diffraction (§ 224). The result was that Young's explanation was not generally accepted; but a modification of his experiment, made by Fresnel, completely settled the matter.

Light, diverging from the point B (Fig. 126), is reflected from two mirrors, OR, OS, which are hinged together at O, and are inclined to each other at a very small angle. After reflection the rays appear to diverge from A and A′, the images of B in OS and OR respectively. Hence A and A′ act as two sources of light, the radiation emitted from each of which is similar in all respects to that emitted from the other. The light has nowhere passed through an aperture, so that the objection made to Young's form of the experiment does not apply, and yet the same effects are observed to occur.

The points A′, A, and B obviously lie on a circle, the centre of which is at O; and the lines OR and OS are respectively perpendi-

cular to A′B and AB. Hence the angle A′BA is equal to the angle of inclination of the mirrors $=\theta$ (say). But $A'OA = 2A'BA = 2\theta$; and

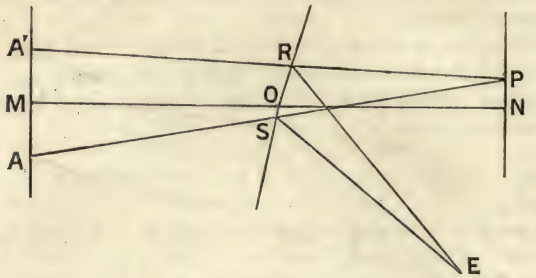


Fig. 126.

OM is practically equal to $OB = r$ (say). Therefore, if we denote ON by r', the formula of last section becomes

$$\frac{2r\theta x}{r+r'} = n\frac{\lambda}{2}.$$

Very accurate adjustments are necessary in order to obtain good results from this form of the experiment.

213. *Lloyd's Experiment.*—Lloyd repeated the above experiment with only one mirror. A ray of light diverging from a slit, A′ (Fig. 127), is reflected in part, at grazing incidence, from a mirror

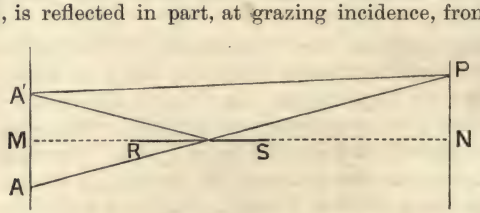


Fig. 127.

RS. We thus obtain two rays of light, one actually diverging from A′, and the other apparently diverging from A, the image of A′ in RS; and these rays produce interference effects as formerly.

Yet one distinct difference is observable. In both forms of the experiment previously described the point N is brightly illuminated, for $AN - A'N = 0$. In Lloyd's experiment N is dark, and the whole system of bright and dark bands is shifted by the breadth of one band. In explanation of this Lloyd suggested that the phase is altered by 180° in the act of reflection.

In all cases the slit through which the light passes should be narrow; but this is not of so much importance in the present case as in the previous cases. For the slit A is the inverted image of A′, and so M is the centre of all corresponding parts of A and A′, the part of A which is nearest to M being the image of the part of A′ which is nearest to M, and so on. Hence the effects of all the parts are strictly superposed at P. But, in the two previous cases, since there is no inversion of A′ with respect to A, the part of A′ which is nearest to M corresponds to the part of A which is farthest from M (M being taken as the middle point of the line joining the central parts of the slits), and so on. Hence the systems of bands due to the light from the various corresponding parts of the two slits are not exactly superposed, and the definition is in consequence less accurate.

214. *Fresnel's Biprism.*—A second form of the experiment, to which also the objection taken to Young's experiment does not apply, is due to Fresnel. RS (Fig. 128) is a glass prism of very obtuse angle. It is placed with its flat face towards M, the source of light. Each half of the prism forms an image of M, so that the rays emerge from

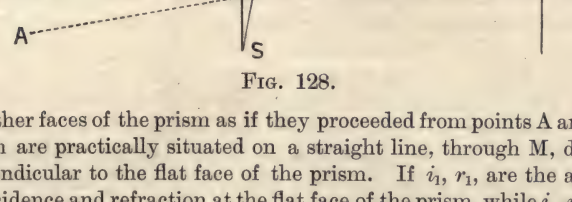


FIG. 128.

the other faces of the prism as if they proceeded from points A and A′, which are practically situated on a straight line, through M, drawn perpendicular to the flat face of the prism. If i_1, r_1, are the angles of incidence and refraction at the flat face of the prism, while i_2, r_2, are the similar angles at the opposite face, the total deviation of the ray MR, *i.e.*, the angle A′RM is (§ 192) $i_1 - r_1 + i_2 - r_2$. These angles being small, i_1 and i_2 are respectively equal to μr_1 and μr_2, μ being the refractive index of the substance of which the prism is composed. Hence the deviation is $(\mu - 1)(r_1 + r_2) = (\mu - 1)\alpha$, where α is the acute angle of the prism. This gives $A'M (= AM) = b(\mu - 1)\alpha$ approximately, b being the distance of M from the biprism, and so the formula giving the value of x becomes

$$\frac{2b(\mu - 1)\alpha x}{b + b'} = n\frac{\lambda}{2},$$

where b' is the distance between the prism and the screen.

215. *Coloured Interference Bands.*—In the immediately preceding sections we have assumed the wave-length to be constant. But the breadth between two adjacent bright or dark bands is proportional to λ, and so the band situated at N is the only one which is colourless. All other bands are coloured, the first red band being about twice as far from N as the first violet band. About a dozen of these bands can be fairly well distinguished when ordinary white light is used; but the succeeding bands of different colours are so superposed that all traces of interference effects practically disappear, and the screen seems uniformly illuminated.

If the quantity a in the formula of § 214 were variable and proportional to λ, x would be constant for all wave-lengths, that is, the bands would be colourless. This effect may be attained by the use of a diffraction grating (§ 233).

In the biprism method the distance between the points A and A′ depends upon the wave-length, and is greater the shorter the wave-length is. The result is that the coloured bands are more widely separated than they usually are.

The introduction of a coloured glass, which diminishes the number of different kinds of light in the interfering beams, produces a very marked increase in the number of bands which are visible. As many as 200,000 bands have been counted when a flame, tinged deeply orange by burning sodium, was employed as the source of light.

When the difference between the lengths of the paths travelled by the two interfering rays is a very large multiple of the wave-length, the nature of the vibrations may have completely altered in the interval of time between the setting out, from the source, of the two waves which simultaneously reach P, so that no interference could occur. But the fact that no more than 200,000 bands have ever been counted does not prove that no more than 200,000 vibrations of the ether at a given point are sufficiently nearly similar to produce continued interference, for we can neither obtain absolutely monochromatic light nor use an infinitely narrow slit. Yet the converse statement, that 200,000 successive vibrations are practically similar, is true.

216. *Displacement of Bands by Refracting Media.*—If a dense medium be placed in the path of one of the two interfering rays, the whole system of bands will be displaced towards that side of MN on which the medium is placed. For if t be the thickness of a medium of refractive index μ, which is traversed by the ray, the effect is the same as if the ray had traversed a thickness, μt, of air. Thus the effective length of the path of that ray is increased by the amount $(\mu - 1)\, t$.

Let L (Fig. 129) represent the medium interposed in the path of the ray A′P. The effective length of A′P is increased, and so the length of AP must be increased. In other words, PN must increase.

Suppose now that L is removed, and that we shift A′ back from the screen through the distance $(\mu-1)t/2$ into the position A_1'. Let also A be moved towards the screen, through the same distance,

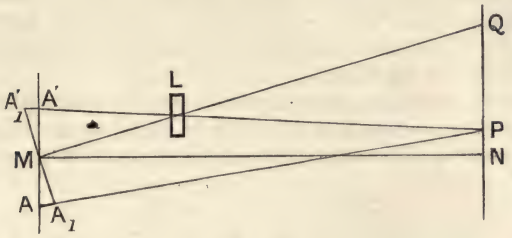


FIG. 129.

into the position A_1. In this way the effective length of A′N is increased, relatively to that of AN, by the amount $(\mu-1)t$; and the central band, originally at N, will now be found at Q, which is such that MQ is perpendicular to A_1A_1'. But $QN/MN = AA_1/AM = (\mu-1)t/2a$. Hence the displacement of the central band (if we are dealing with monochromatic light) is

$$QN = \frac{b(\mu-1)t}{2a}.$$

When the light is not monochromatic the displacement of the central (which is then the brightest) band could only be given by this formula if the refracting substance did not produce dispersion, *i.e.*, if μ were independent of λ. The brightest effect will really be produced at a place where the rate of variation of QN with λ is a minimum, for, at such a place, the various adjacent coloured bands are most nearly superposed.

By means of the formula the refractive index of the interposed substance may be found with extreme accuracy. The method is specially applicable to the determination of the refractive indices of gases.

217. *Interference Bands in Spectra.*—If a beam of white light, diverging from a narrow slit, be made parallel by a suitable lens, and then be refracted by a prism, the usual continuous spectrum will be obtained. But if a plate of a refracting substance be interposed in the path of one half of the beam, the spectrum will be

crossed by dark bands. The reason is that one half of the rays are retarded relatively to the other half, and so interference effects are produced. Those rays, the relative retardation of which amounts to a semi-wave-length, are obliterated.

Various forms of this experiment are described by Powell, Fox, Talbot, Brewster, and Stokes.

218. *Colours of Thin Plates. Reflected Light.*—Thin films of transparent substances are frequently observed to be brilliantly coloured. The colours vary with the angle of incidence and with the thickness of the film.

Familiar examples occur in the cases of a soap bubble, of the wing of the common house fly, and of highly tempered steel, etc. In the latter case the thin film consists of an oxide formed on the

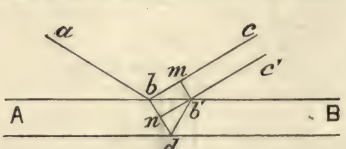


FIG. 130.

surface of the steel at a high temperature. Very old glass vessels frequently exhibit these colours from the partial splitting away of thin films at the surface of the glass.

The wave theory gives a complete explanation of these phenomena.

Let AB represent a film, of (small) thickness t, of a substance the refractive index of which is μ. A ray, ab, falling upon the upper surface of the plate is partially reflected along bc and in part is refracted along bd. The refracted ray suffers partial reflection in the direction db' and finally emefges from the substance in the direction $b'c'$ parallel to bc.

If perpendiculars $b'm$ and $b'n$ be dropped from b' upon bc and bd, the parts bm, bn of these paths intercepted between b and the feet of the perpendiculars are described in equal times. Hence the effective difference of path described by the two rays is $nd+db'$, which is equal to $2t \cos r$, where r is the angle of refraction. And this portion is described in a substance of refractive index μ, so that the equivalent path in air is

$$2\mu t \cos r.$$

It might, therefore, be expected that the effects of the two rays would be mutually intensified when this quantity is an integral

multiple of a wave-length, that they would mutually annul each other when it is an odd multiple of a semi-wave-length, and that, when the thickness of the plate is much smaller than a semi-wave-length of violet light, all rays would be intensified, so that white light would be reflected.

But the exact reverse of these effects are observed. When the thickness of the plate is very small no light is reflected, and, when the quantity $2\mu t \cos r$ is an odd multiple of half a wave-length, the light is strongly reflected. These results are precisely those which would occur if, in the acts of reflection at the upper and under faces, a difference of phase of half a period were introduced.

The conditions under which the two reflections take place are exactly opposed to each other. In the one case the light is passing from a rarer into a denser medium : in the other it is passing from a denser into a rarer medium. Hence, reasoning by analogy from the effects of impact of two elastic balls of different masses, Young pointed out that the relative acceleration of phase which seems to be required ought to be produced. [The propagation of waves along a rope composed of two parts of different linear densities, is precisely analogous. A wave propagated along the less dense portion is in part reflected from the junction with a complete reversal of phase. (As an extreme case imagine the rope to be fixed at the junction. This corresponds to infinite density of the second part.) A wave travelling along the more dense portion is partly reflected at the junction without change of phase.]

Young pointed out that if his explanation were correct an entire reversal of the effects should occur when the reflecting plate was intermediate in density between the media on either side of it. Further, he carried out such an experiment, and found that his prediction was verified. Lloyd's experiment (§ 213) furnishes another verification of the correctness of Young's explanation.

The effective difference of path, $2\mu t \cos r$, decreases as the angle of incidence increases, and therefore the wave-length of the reflected light decreases as the angle of incidence increases. If the refractive index and the thickness of the plate be sufficiently large the series of colours may be repeated a number of times, but, if ordinary white light be used, partial overlapping will occur between all the series above the second, for the wave-length of the extreme red light of the spectrum is approximately double of that of the extreme violet light.

219. The above explanation of the reflection of light from thin plates is not quite complete. The intensity of the reflected ray, bc, is always greater than that of the ray $b'c'$, and so complete

annulment of light is not accounted for. But complete annulment does take place. A complete treatment of the problem was given by Poisson, who pointed out that all the various rays which emerge at b' must be taken into account. The ray which enters at b and suffers one internal reflection at d before it passes out of the plate at b'

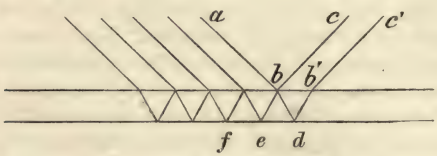


FIG. 131.

has the greatest effect in producing the final result; but the ray which suffers two such internal reflections (at e and d) before emergence also has a considerable effect. Similarly, those which have undergone three, four, etc., internal reflections, have each an appreciable, though rapidly diminishing, share in the ultimate result.

The effective difference of path between the ray which has suffered n such internal reflections, and the ray which is once reflected externally at b is $2n\mu t \cos r + \lambda/2$, the semi-wave-length being added in order to take account of the acceleration of phase produced in the act of reflection at b. This being taken into consideration it is found that the intensity of the light reflected from the plate *does* vanish when $2\mu t \cos r$ is an even multiple of $\lambda/2$, and that it is a maximum when $2\mu t \cos r$ is an odd multiple of $\lambda/2$.

220. *Colours of Thin Plates. Transmitted Light.*—The light which is transmitted through the plate is complementary to that which is reflected from it; that is, the kinds of light which are absent from the reflected beam are precisely those which are present in the transmitted beam.

The intensity of the reflected light is never equal to that of the incident light, and so the intensity of the transmitted beam never entirely vanishes. Also, since the minimum intensity of the reflected light is zero, the maximum intensity of the transmitted light is equal to the intensity of the incident light.

221. *Newton's Rings.*—Newton observed the colours produced by the interference of rays reflected from both sides of a thin film of air enclosed between two pieces of glass. One of the pieces of glass had a plane surface; the surface of the other was convex and spherical.

The thickness of the film of air at a distance, d, from the point of contact is approximately $d^2/2R$ where R is the radius of the spherical

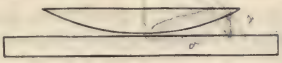


FIG. 132.

surface. Hence the condition that light of wave-length λ shall be intensified is

$$\frac{\mu' d^2}{R} \cos r = n\lambda + \frac{\lambda}{2} + = (2n+1)\frac{\lambda}{2}$$

n being any integer, and so the point of contact is surrounded by a series of bright rings. In this formula, r is the angle of refraction from glass into air and μ' is the reciprocal of the refractive index of glass. The radii of successive bright rings are therefore given by

$$d = \sqrt{\mu R \sec r\ (2n+1)/2\ .\ \lambda},$$

where μ is the refractive index of the glass referred to air.

It follows from this formula that the replacement of the film of air by a denser substance would cause all the rings to close in somewhat towards their common centre. This result is proved by experiment, and hence we get another proof of the fact that light travels slower through a medium such as water than it does through air.

The successive radii are proportional to the square roots of the natural numbers—of the even numbers in the case of the dark rings, and of the odd numbers in the case of the bright rings, and so successive rings enclose equal areas.

The radii also increase as the wave-length increases, and so the first red ring is farther from the centre than the first blue one is.

Lastly, d increases when the angle of incidence increases.

When the two pieces of glass are pressed sufficiently close together a black spot appears at the centre. The central thickness is then very small in comparison with the wave-length of any visible light, and so the reflected light vanishes; for the effective length of the paths traversed by rays which emerge at a given point after internal reflection is practically the same as that of the light which is directly reflected at the same point without entering the thin film, and so the two sets of rays practically differ in phase by half a period.

The transmitted light is complementary to that which is reflected. The central portion is therefore white.

Theory indicates that if the refractive index of the film be intermediate between the indices of the two transparent media which bound it, the rings seen by reflection should commence from a white centre. Young verified this prediction by means of a film of oil of sassafras enclosed between a lens of crown glass and a lens of flint glass.

222. *Colours of Mixed Plates.*—If a bright object be viewed through an intimate mixture of two media of different refractive indices (*e.g.*, a mixture of oil and air enclosed between glass plates), colours are observed to which Young gave the name of 'colours of mixed plates.' The colours are arranged in rings precisely as in the case of those seen by transmission of light through a thin homogeneous plate, but the whole system is on a larger scale. The phenomenon is due to the interference of the rays which pass through the different media and so suffer relative change of phase.

When the incident light is oblique and a dark object is placed behind the plates, the system resembles that which is ordinarily seen by reflection, for one of the interfering portions is reflected and undergoes the usual acceleration of phase.

223. *Colours of Thick Plates.*—Brewster observed that, in certain circumstances, interference may be produced by means of plates the thickness of which is not small in comparison with the wave-length of light.

AB and BC (Fig. 133) represent two such plates of parallel glass, which are precisely equal in thickness, and are inclined to each other at a small angle, α.

A pencil of light, Pm, falls perpendicularly upon the plate BC, and, passing through it, is partly reflected from the first surface of

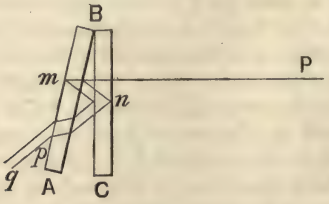


Fig. 133.

AB and in part is refracted into the plate AB. A portion of the refracted part is reflected at m. If r be the angle of refraction in AB the effective difference of path so produced between the two

portions of light is $2\mu t \cos r = 2\mu t \cos \sin^{-1}(1/\mu \,.\, \sin \alpha)$, where μ is the refractive index and t is the thickness of the glass. A similar action occurs at the plate BC, and the rays which were reflected from the first surface of AB sustain, relatively to the other rays, an effective increase of path to the amount $2\mu t \cos r' = 2\mu t \cos \sin^{-1}(1/\mu \,.\, \sin 2\alpha)$, r' being the angle of refraction in BC. The effective difference of path of the two rays, pq, which finally emerge from the side of AB remote from P, is therefore

$$2\mu t\Big(\cos \sin^{-1}(\frac{1}{\mu} \sin \alpha) - \cos \sin^{-1}(\frac{1}{\mu} \sin 2\alpha)\Big).$$

Interference occurs when this quantity is sufficiently small.

Jamin has applied this principle to the construction of an extremely sensitive instrument for the measurement of refractive indices.

Newton observed interference effects when he allowed light to fall upon the surface of a concave glass mirror which was silvered behind. The mirror was everywhere of uniform thickness and the light was admitted through a small opening in a sheet of white paper—the opening being situated at the centre of curvature of the mirror. A few broad coloured rings, resembling those due to light transmitted through a thin plate, were seen on the paper. All these rings were concentric with the opening through which the light passed.

The origin of these colours is totally different from that of the colours which Brewster observed. The rings are due to the interference of light, ordinarily reflected at the silvered surface of the mirror and then scattered by particles of dust on the first surface, with light, also reflected from the silvered surface, but which had been previously scattered (or, rather, *diffracted*) by particles of dust upon the first surface of the mirror.

When the mirror is slightly inclined the centre of the coloured rings is situated midway between the opening in the paper and the image of it which is formed upon the paper. This central spot is alternately bright and dark (when homogeneous light is used), as the distance between the opening and its image increases; it undergoes a rapid variation of colour when the incident light is white.

224. *Diffraction.*—The principle by means of which Huyghens explained the rectilinear propagation of light has already been given (§ 186). The following remarks of Stokes on this subject are specially worthy of note.

'When light is incident on a small aperture in a screen, the illumination at any point in front of the screen is determined, on the undulatory theory, in the following manner. The incident waves are conceived to be broken up on arriving at the aperture; each element of the aperture is considered as the centre of an elementary disturbance, which diverges spherically in all directions, with an intensity which does not vary rapidly from one direction to another in the neighbourhood of the normal to the primary wave, and the disturbance at any point is found by taking the aggregate of the disturbances due to all the secondary waves, the phase of vibration of each being retarded by a quantity corresponding to the distance from its centre to the point where the disturbance is sought. The square of the co-efficient of vibration is then taken as a measure of the intensity of illumination. Let us consider for a moment the hypothesis on which this process rests. In the first place it is no hypothesis that we may conceive the waves broken up on arriving at the aperture: it is a necessary consequence of the dynamical principle of the superposition of small motions, and if this principle be inapplicable to light, the undulatory theory is upset from its very foundations. The mathematical resolution of a wave, or any portion of a wave, into elementary disturbances must not be confounded with a physical breaking up of the wave, with which it has no more to do than the divisions of a rod of variable density into differential elements, for the purpose of finding its centre of gravity, has to do with breaking the rod in pieces. It *is* a hypothesis that we may find the disturbance in front of the aperture by merely taking the aggregate of the distubances due to all the secondary waves, each secondary wave proceeding as if the screen were away; in other words, that the effect of the screen is *merely to stop* a certain portion of the incident light. This hypothesis, exceedingly probable, *à priori*, when we are only concerned with points at no great distance from the normal to the primary wave, is confirmed by experiment, which shows that the same appearances are presented, with a given aperture, whatever be the nature of the screen in which the aperture is pierced; whether, for example, it consist of paper or foil, whether a small aperture be divided by a hair or by a wire of equal thickness. It is a hypothesis, again, that the intensity in a secondary wave is nearly constant, at a given distance from the centre, in different directions very near the normal to the primary wave; but it seems to me almost impossible to conceive a mechanical theory which would not lead to this result. It is evident that the difference of phase of the various secondary waves which agitate a given point must be determined by the difference of their radii, and if it should

afterwards be found necessary to add a constant to all the phases the results will not be at all affected. Lastly, good reasons may be assigned why the intensity should be measured by the square of the co-efficient of vibration.'

225. Huyghens' construction, if rigorously carried out, would indicate the existence of a wave running back towards the source as well as a wave which travels forwards. Analogy points to the conclusion that the part of the construction which leads to a reverse wave must be ignored. For example, the investigation of § 73 shows that no wave can travel backwards from a disturbance which runs along a stretched cord. But Stokes, in his paper on the Dynamical Theory of Diffraction, of the introduction to which the above quotation forms part, has shown from purely dynamical principles, that the disturbance in a secondary wavelet is a maximum in the direction of the wave-normal, and that it diminishes constantly as the direction considered is inclined more and more to the normal, ultimately becoming zero in the direction opposite to that in which the primary wave travels. He then shows that the result of the superposition of all the secondary effects is the same as if the wave (assumed to be practically plane, *i.e.*, of radius which is large in comparison with the wave-length, a condition always satisfied in experiment) had not been supposed to be broken up into a series of separate centres of disturbance, and that no back-wave is propagated.

226. *Effect of a Rectilinear Wave.*—We have already stated that Fresnel showed that Huyghens' principle, according to which the new wave front is found to be the envelope of the secondary wave-fronts, should be explicitly associated with the principle of interference if it is to give a complete explanation of the rectilinear propagation of light. The envelope is the locus of points each of which is simultaneously reached by more than one secondary disturbance the phases of which are identical. It is, therefore, the locus of points at which the light has great intensity.

The necessity for the introduction of the principle of interference will appear very evidently from the following investigation of the effect of a rectilinear wave at any external point.

Let AB (Fig. 134) represent a portion of a linear wave which extends to infinity in both directions, and let P be the point at which we have to determine the effect of the wave. Draw PM perpendicular to AB and take points m, m', m'', etc., such that $Pm - PM = Pm' - Pm = Pm'' - Pm' = \text{etc.}, = \lambda/2$, where λ is the wave-length of the light emitted from the various points of AB.

The length of the half-period element Mm is $\sqrt{(a+\lambda/2)^2 - a^2}$

where a is the length of PM. When λ is so small in comparison with the other length involved that it may be neglected, this becomes $\sqrt{a\lambda}$. Similarly Mm', Mm'', etc., are respectively equal to

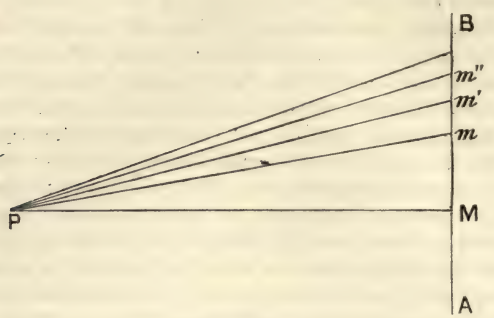


FIG. 134.

$\sqrt{2a\lambda}$, $\sqrt{3a\lambda}$, etc. Hence the lengths of the successive half-period elements, from M outwards, are $\sqrt{a\lambda}$, $\sqrt{a\lambda}\,(\sqrt{2}-1)$, $\sqrt{a\lambda}\,(\sqrt{3}-\sqrt{2})$, etc., and the limit to which they ultimately approach is $\lambda/2$.

If we divide each element into the same number of infinitesimal portions, the light sent out by the first portion of the first element differs in phase from that emitted by the first portion of the second element by one half of a period. Similarly, the light emitted by the second portion of the first element differs in phase by one half of a period from that emitted by the second portion of the second element, and so on. Now the effects at P of the various parts of the first element are not quite compensated by the effects of the corresponding parts of the second element. For the breadth of the parts of the first element is rather greater than that of the parts of the second; and the inclination of each part to the line joining it to P, and also its distance from P, increase as the part is more remote from M. But the difference between the effects of the corresponding portions of the n^{th} and the $n+1^{th}$ elements is vanishingly small when n is large.

Now, as λ is a very small length it follows that a very large number of half-period elements are included in a small portion of AB in the near neighbourhood of M, and, consequently, only a small part of the wave near M produces any effect at P. Hence, a small opaque object placed on the line PM would entirely prevent the wave AB from producing any effect at P.

Hence the propagation of light is practically rectilinear when λ is

so small that its square may be neglected in comparison with the other quantities involved.

Let e_2, e_2, etc., be the effects at P of the first, second, etc., half-period elements of MB. The total effect (taking account also of the portion MA) is

$$2(e_1 - e_2 + e_3 - e_4 + \ldots\ldots - e_{2s} + \ldots\ldots) .$$

These various terms are in descending order of magnitude, and therefore it appears that the total effect is smaller than the effect of the first half-period portions at M.

The difference between any two successive terms is small in comparison with the magnitude of either, and so, writing the above expression in the form

$$e_1 + (e_1 - e_2) - (e_2 - e_3) + (e_3 - e_4) + \ldots\ldots ,$$

we see that the total effect at P is approximately equal to the effect of one half-period element at M.

227. *Effects of Plane and Spherical Waves.*—Let the plane of the paper represent a plane wave, the effect of which at a point, P, is to be found, and let M (Fig. 135) be the foot of the perpendicular drawn from P to the plane.

FIG. 135.

From P as centre describe successive spheres of radii $MP + \lambda/2$, $MP + 2\lambda/2$, etc. The spheres will divide the plane wave into concentric zones, called half-period zones, or, sometimes, Huyghens' zones.

Dividing each of these zones into the same number of infinitesimal annular portions, we observe that the effect of each portion of one zone is nearly annulled by the effect of the corresponding portion of the succeeding zone, and that the annulment is practically complete at a short distance from the point M—precisely as in the similar investigation of last section. Hence the effect produced at

P is that due to a few half-period zones in the neighbourhood of the wave-normal which passes through P, and is practically equal to half the effect of the first zone.

Let AMB (Fig. 136) represent a spherical wave diverging from O.

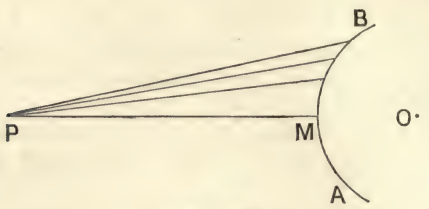


FIG. 136.

To find the effect at P we must, as above, divide AMB into half-period zones surrounding M the point in which OP intersects AB.

Reasoning similar to the foregoing shows that the effect of the wave at P is equivalent to half of that produced by the first zone.

Now it is easy to see that the phase of the vibration due to the first zone differs from that due to the secondary wave at M by one quarter of a period. For if OM be large in comparison with the wave-length—a condition which is satisfied in all experimental observations—the first zone is practically plane. And, further, if it be broken up into $2n$ infinitesimal rings, of equal area, surrounding the point M as centre, the amplitudes of the vibrations produced at P by each of these annular portions will be practically equal to one another. Hence (§ 52) the phase of the resultant of the effects of the 1^{st} and the $2n^{th}$ annuli is halfway between those of its components. This is true also of the phase of the resultant of the effects of the 2^{nd} and the $(2n-1)^{th}$ annuli, and so on. But the phase of the vibration at P due to each annulus varies uniformly from the 1^{st} to the $2n^{th}$ annulus. Therefore the phase of the resultant vibration at P due to the complete zone is one-quarter of a period behind that due to the vibration at M. The same statement must be true of the vibration produced by the whole wave if it agrees in phase with that produced by the first zone.

228. *Diffraction at a Straight Edge.*—We are now in a position to determine the effects produced by any given portions of a wave which diverges from a luminous point—the remaining portions being intercepted by opaque obstacles. This involves the carrying of our investigations beyond the stage in which the wave-length may be assumed to be small in comparison with all other quantities

involved. We shall find that, under this new condition, light is no longer propagated in straight lines, but is bent, or *diffracted*, into the geometrical shadows precisely as sound is.

AMB (Fig. 137) represents a spherical wave, which, diverging from the point O, is partially intercepted by an opaque object MN. We have to determine (1) the effect at any point P outside the geo-

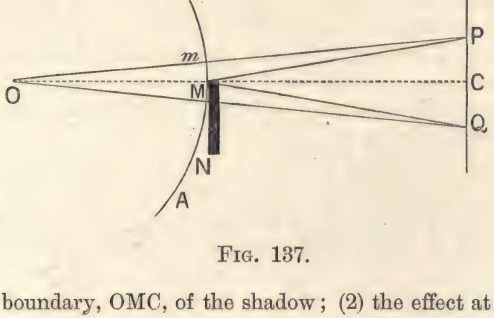


FIG. 137.

metrical boundary, OMC, of the shadow; (2) the effect at any point inside Q the geometrical shadow.

Join OP and MP, and let OP meet AB in m.

When mM contains a considerable number of half-period elements, the wave produces practically its full effect at P. Let e_1, e_2, etc., be the effects of the first, second, etc., half-period elements in the neighbourhood of m; and let E be the effect of the semi-wave mB. The effect at P is E, $E+e_1$, $E+e_1-e_2$, etc., according as the number of elements included in mM is O, 1, 2, etc. Hence the effect at P is a maximum, or a minimum, according as mM contains an odd, or an even, number of half-period elements; that is, according as, in the formula

$$MP - mP = n\frac{\lambda}{2},$$

n is odd or even. When n and λ are given, the locus of P is a hyperbola the foci of which are O and M. Now, if we denote PC by x, OM by a, and MC by b, we get $OP^2=(a+b)^2+x^2$ and $\overline{MP}^2=b^2+x^2$. These expressions give approximately $OP=a+b+x^2/2(a+b)$, and $MP=b+x^2/2b$. Hence the above formula becomes

$$x^2\frac{a}{2b(a+b)}=n\frac{\lambda}{2}.$$

At a point Q, within the geometrical shadow, the most effective

portions of the wave are intercepted by the obstacle. The effect at Q is practically $\frac{1}{2}e_2$, $\frac{1}{2}e_3$, etc., according as MN intercepts 1, 2, etc., of the most powerful elements. Hence the illumination inside the geometrical shadow dies away as the distance of the illuminated point from the geometrical boundary increases.

Diffraction fringes, resembling those just described, appear outside the geometrical shadow on both sides of a narrow obstacle, such as a thin wire or a hair. But, in addition, a series of finer bands, of constant breadth, make their appearance inside the geometrical shadow if the obstacle is sufficiently narrow. These are caused by interference of the light diffracted at both sides of the obstacle, for, as we have seen, the effect of each unintercepted portion of the wave is practically the same as that of a luminous line placed close to the straight edge of the obstacle.

229. *Diffraction at a Narrow Slit.*—Let MN (Fig. 138) represent a narrow opening in an opaque obstacle, and let a wave AB diverge from a point O, which is situated on the line drawn from the middle point of MN at right angles to the plane of the obstacle.

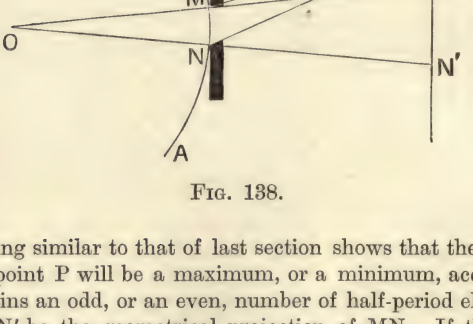


FIG. 138.

Reasoning similar to that of last section shows that the illumination at a point P will be a maximum, or a minimum, according as MN contains an odd, or an even, number of half-period elements.

Let M′N′ be the geometrical projection of MN. If the screen, PM′N′, be so far from MN that NM′ − MM′ (or MN′ − NN′) is less than a semi-wave-length, a fringe of alternately bright and dark bands will appear on each side of the geometrical projection of MN. But if the distance between the obstacle and the screen be so small that NM′ − MM′ is greater than a semi-wave-length, bands will appear between M′ and N′.

230. *Diffraction at a Circular Aperture. Zone Plates.*—Draw

the line OP (Fig. 139) through the centre of the aperture and perpendicular to its plane. We shall determine the general effect at P of light diverging from O.

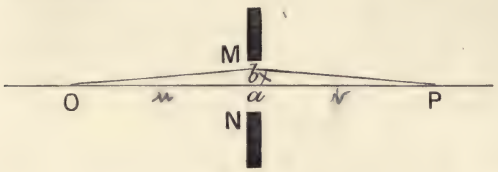


FIG. 139.

The illumination at P is a maximum, or a minimum, according as MN contains an odd, or an even, number of half-period zones.

Let a be the centre of the aperture, and let b be the outer edge of the n^{th} zone. Denote the lengths of Oa, $a\text{P}$, and ab by u, v, and x respectively. We get approximately

$$Ob = u + \frac{x^2}{2u},$$

and

$$b\text{P} = v + \frac{x^2}{2v};$$

therefore

$$Ob + b\text{P} = u + v + \frac{x^2}{2a} + \frac{x^2}{2b}.$$

Hence

$$Ob + b\text{P} - \text{OP} = \frac{x^2}{2}\left(\frac{1}{u} + \frac{1}{v}\right).$$

But this length is equal to $n\lambda/2$, and so

$$x^2\left(\frac{1}{u} + \frac{1}{v}\right) = n\lambda.$$

Thus the consecutive values of x are proportional to the square roots of the natural numbers.

As P approaches MN, the number of half-period zones in the aperture increase, and so the illumination at P passes through a succession of maxima and minima. The various points at which the maxima and minima occur are given by the expression

$$v = \frac{ur^2}{un\lambda - r^2}$$

where r is the radius of the aperture.

From the way in which λ is involved, it is evident that the position of P corresponding to maximum illumination approaches nearer to the aperture when the wave-length increases; so that an eye, which advances to the aperture along Pa, will perceive a rapid periodic variation in the colour of the light which reaches it.

If, as formerly (§ 228), we denote by e_1, e_2, etc., the effects of the light which passes through the successive half-period-zones, the total effect is

$$e_1 - e_2 + e_3 - e_4 + \ldots\ldots$$

The effect will therefore be much greater than it otherwise could be if the even zones be made opaque. Such an arrangement constitutes a *zone plate*. If n be the number of zones (alternately open and opaque) in a zone plate of radius r, the formula

$$\frac{1}{u} + \frac{1}{v} = \frac{n\lambda}{r^2}$$

shows (§ 193) that the plate acts as a condensing lens, the principal focal length of which is $r^2/n\lambda$. But, in the case of the zone plate, all rays do not take the same time to pass between the conjugate foci; and, further, the focus for red rays is nearer to the plate than the focus for blue rays is. In these points it differs from a lens.

231. *Diffraction at an Opaque Disc.*—A point at the centre of the geometrical shadow is almost as brightly illuminated as if the disc were removed. If this disc removes $n-1$ half-period zones, the effect of the remaining zones is practically equal to one half of that of the n^{th} zone. But, so long as n is not large, the effect of the n^{th} disc is not greatly different from that of the first. This theoretical result was first pointed out by Poisson, and was verified experimentally by Arago.

232. *Coronæ. Young's Eriometer.*—If a number of very small and nearly equal particles be closely distributed in the space intervening between a luminous object and the eye, the object will appear to be surrounded by luminous rings. These are due to diffraction of the light which passes the edges of the particles. The *coronæ*, which are sometimes seen surrounding the sun or the moon, are caused by the presence of small globules of water in the atmosphere. They are coloured blue inside, red outside, and increase in size when the diameters of the globules diminish. [If, therefore, the coronæ are observed to contract, the moisture in the atmosphere is condensing, and rain may be expected to follow; conversely, if the rings dilate, dry weather will in general ensue.]

Young's *Eriometer* was devised for the purpose of measuring the

diameters of small objects. It consists of a metal plate, in which a small hole is drilled. The plate is also perforated by a circle of still smaller holes which surround the large hole as a centre. A flame is placed behind this plate, and the light which passes through the holes is examined through glass plates which contain between them the (equal) particles the size of which is to be determined. The large opening in the metal plate is surrounded by coloured rings, and the distance between the metal plate and the glass plates is altered until any one particular ring coincides with the circle of small holes. This distance varies inversely as the radius of the ring, which itself, as we have just seen, varies inversely as the diameter of the particles. One experiment, in which the diameter of the particles is known, and the distance between the plates is measured, is sufficient to enable us to calculate the unknown diameters of other sets of particles.

233. *Diffraction Gratings.*—A diffraction grating may consist of a glass plate, upon which a great number of extremely fine equidistant parallel lines are ruled by means of a diamond point. The grooves are practically opaque, for light incident on them is reflected back in all directions. On the other hand, the glass between the grooves is transparent, and the light which passes through is diffracted in all directions.

Let AB (Fig. 140) represent a (highly-magnified) portion of the grating, the dark parts indicating the grooves, and the light parts indicating the intervening spaces; and let aP and bP represent the

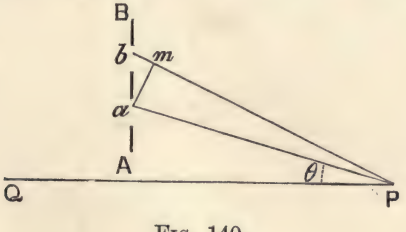


FIG. 140.

paths of rays which reach P from similarly-situated parts of the openings a and b. From a drop a perpendicular am upon bP. The distance ab being very small in comparison with the distance of P from the grating, bm is practically equal to bP$-a$P.

Now suppose that parallel rays from a narrow slit parallel to the grooves fall perpendicularly upon the grating. An eye placed at P will see the slit through the grating in the direction PQ. The

angle bam is equal to the angle $a\mathrm{PQ}(=\theta$, say); and so $bm = ab \sin\theta$. Therefore a maximum or minimum effect will be produced at P according as n is even or odd in the expression

$$ab \sin\theta = n\lambda/2.$$

The length ab is known, since it is the reciprocal of the number of grooves ruled in unit breadth of the grating.

If monochromatic light be used, a series of coloured images of the slit will be seen at different angular distances from the line PQ. If white light be used, a series of spectra will be observed, in each of which the violet light is less bent from its original direction than the red light is. The spectra are said to be of the first, second, etc., order, according as n has the values 1, 2, etc. All the spectra beyond the second partially overlap each other.

Very accurate measurements of wave-length may be made by means of the grating; and the spectra obtained from all gratings are identical, except as regards *scale;* that is, there is no trace of irrationality in the dispersion (§ 197). And, further, if θ is nearly zero, the dispersion between any two rays is practically proportional to the difference of their wave-lengths. This condition may be attained by inclining the grating to the direction of the incident light at a suitable angle. The spectrum thus produced is called a *normal spectrum.*

Diffraction spectra may be obtained by reflection from a ruled metallic surface. Rowland's concave gratings are ruled on the polished surface of a portion of a cylinder of speculum metal.

CHAPTER XIX.

DOUBLE REFRACTION. POLARISATION.

234. *Double Refraction.*—In our consideration of the refraction of light, we have hitherto dealt only with those cases in which a single refracted ray occurs.

Bartholinus, in 1669, described the phenomenon of double refraction as observed by him in Iceland spar.

A single ray of light incident upon the surface of Iceland spar in general gives rise to two refracted rays. One of these obeys the ordinary law of refraction, but the other follows a totally different law. The former is called the ordinary, and the latter the extraordinary, ray.

All crystalline minerals, except those belonging to the cubic system, possess the property of double refraction.

The fundamental form in which Iceland spar crystallizes is the rhombohedron. The angles of the faces are either acute or obtuse. The obtuse angles are all equal to each other, and the acute angles

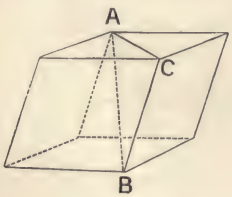


FIG. 141.

are all equal also. Two of the solid angles (A and B, Fig. 141) of the rhombohedron are bounded by three obtuse angles. All other angles, such as C, are bounded by one obtuse and two acute angles. The *axis of the crystal* is a line which is equally inclined to the three edges meeting at an obtuse-angled corner. If we make all the edges of the block equal in length, the crystalline axis will be AB,

the diagonal joining the two obtuse-angled corners. A plane ACB, which passes through the crystalline axis and the shorter diagonal AC of a rhombic face, is called a *principal section* of the crystal.

In Iceland spar, and in many other crystalline substances, all the optical properties are symmetrical about the axis of form. Any *direction* in such a substance, which is parallel to the axis of form, is, therefore, called the *optic axis;* and all such substances are called *uniaxal* crystals.

If the spar be cut by a plane in any direction, and a ray of light falls upon the surface so formed, both an ordinary and an extraordinary ray will in general be produced; and, in most cases, the latter will not lie in the plane of incidence. But if the plane be perpendicular to the optic axis, both rays coincide if the incidence is normal. This also occurs if the optic axis lies in the refracting surface, and the incidence is normal; and, further, in this case the extraordinary ray obeys the ordinary law so long as the plane of incidence is perpendicular to the optic axis.

These various phenomena were investigated very fully by Huyghens, and he was led to adopt a construction for the wave-front in the interior of the crystal which he himself proved experimentally to accord very accurately with the observed facts. More severe tests of his construction were made by Wollaston in 1802; and recently Stokes, Mascart, and Glazebrook have verified its accuracy to the full extent attainable by modern methods of measurement.

235. *Huyghens' Construction.*—Huyghens had previously explained the propagation of light in homogeneous isotropic media by

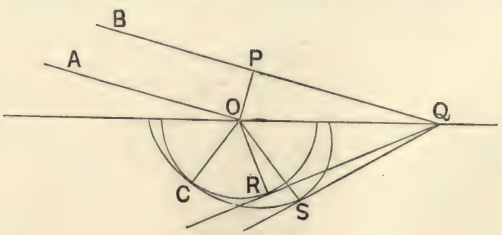


FIG. 142.

the assumption that the wave-surface was spherical (§§ 186, 200). To explain double refraction in uniaxal crystals, he assumed that the wave-surface consists of an ellipsoid of revolution the axis of symmetry of which is coincident with the optic axis, and a sphere which touches the ellipsoid at the extremity of its axis of symmetry.

The spherical portion of the surface corresponds to uniform speed of propagation in all directions; and the incident ray being given, we can determine from it, by the method of § 200, the direction of the ordinarily refracted ray. The ellipsoid indicates unequal speed of propagation in different directions, and from it we can determine the direction of the extraordinarily refracted ray by a similar process.

Let O (Fig. 142) be the point at which an incident ray AO meets the surface OQ, and let BPQ be another ray parallel to AO, so that OP, which is perpendicular to both, may represent a portion of a plane wave-front. In the time in which light moves from P to Q, the ordinary ray will have passed over a distance OR, such that $PQ = \mu OR$, where μ is the ordinary index of refraction. A plane through Q, perpendicular to the plane of incidence, will touch a sphere drawn from O with radius OR in a point R, and OR is the direction of the ordinary ray. The plane RQ is the ordinarily refracted wave-front.

If OC is the optic axis, the radii of an ellipsoid CS, which has OC as its semi-diameter of revolution, will represent the speeds of propagation of the extraordinary ray in different directions. A plane passing through Q, and perpendicular to the plane of incidence, will touch CS in a point S such that OS is the direction of the extraordinary ray; and the ratio PQ/OS is equal to μ', the index of refraction for all extraordinary rays which pass through the crystal in the direction OS.

In Iceland spar, OC is the shortest radius of the ellipsoid; in quartz it is the largest radius. All crystals which resemble Iceland spar in this respect are called *negative* crystals; those which resemble quartz are called *positive* crystals. In the former, the extraordinary index is less than the ordinary; in the latter, the reverse is the case.

If, in this figure, the point C lies out of the plane of the paper, the point S will in general lie outside it also; that is, the extraordinary ray will not be in the plane of incidence. This will be so even if the incidence is perpendicular; for the new wave-front will be a plane parallel to OQ, and this will in general touch CS in a point which does not lie in the plane of the paper.

236. *Special Sections of the Surface.*—(1) Let the refracting surface be perpendicular to the optic axis (Fig. 143). At normal incidence there is no separation of the two rays; but, as the angle of incidence increases, the extraordinary ray separates out farther from the normal than the ordinary one does. If the plane of incidence be rotated around OC, the two rays each maintain a

fixed inclination so long as the angle of incidence remains constant.

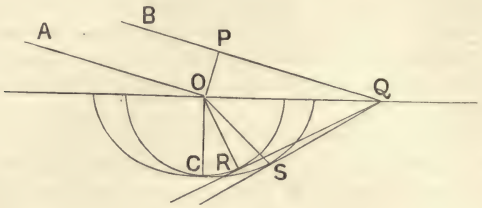


FIG. 143.

(2) Let the refracting surface and the plane of incidence intersect in the optic axis (Fig. 144). At normal incidence there will be no separation of the two rays as regards direction, though the extraordinary ray will travel with greater speed than will the ordinary

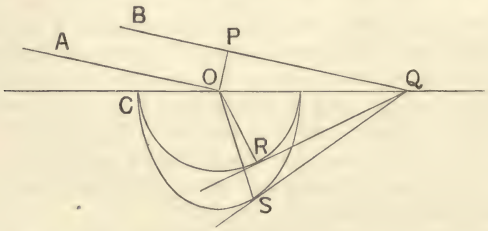


FIG. 144.

ray. And, when the angle of incidence increases, the former does not separate out so far from the normal as the latter does; for, from the properties of the ellipse and circle with a common diameter, R and S lie on a line which is perpendicular to O Q.

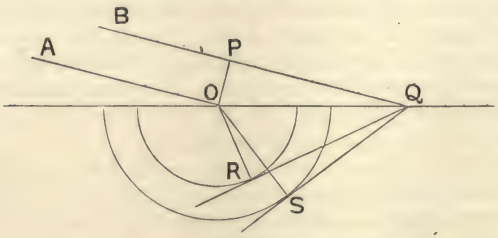


FIG. 145.

(3) Let the refracting surface contain the optic axis, while the plane of incidence is perpendicular to it (Fig. 145). The section of the ellip-

soid becomes a circle, and so the extraordinary ray obeys the ordinary law, though its index of refraction is less than that of the ordinary ray. At normal incidence, this case becomes identical with the last.

237. *Polarisation.*—Huyghens observed that the intensities of the two beams produced by refraction in a block of Iceland spar are equal. And he further noticed that each of these beams was in general subdivided into two others, of *unequal* intensity, on transmission through a second block.

When the principal sections of the two blocks are parallel, no more than two beams are produced: the ordinary ray in the first block passes through the second, without change of direction, as an ordinary ray; and the extraordinary ray passes through also without any change. And, when the principal sections of the blocks are at right angles to each other, two rays only are transmitted; but the ordinary ray in the first block passes through the second as an extraordinary ray, while the extraordinary ray in the first becomes an ordinary ray in the second. In all other relative positions of the two principal sections, each ray in the first is subdivided into two in the second. As the second block is turned round from the position in which its principal section was parallel to that of the first, the two original beams gradually diminish in intensity as the intensities of the newly-produced beams increase. When the principal sections are inclined at an angle of 45° to each other, all the four rays are equally intense. The changes then proceed in the same direction until the inclination of the principal sections is 90°, when the original beams vanish; after this, if the inclination be still further increased, the changes proceed in the reverse order until, at 180°, the beams again pass unchanged through the second block.

Huyghens remarked that the rays which had passed through the first block seemed to have acquired some *form or disposition* which led to the production of these phenomena. Newton spoke of them as possessing *sides*. But it was not until more than a century afterwards that a complete explanation was found as the result of an accidental discovery.

Malus, happening to examine through a doubly refracting prism the light reflected from the windows of the Luxembourg Palace, observed that each ray alternately disappeared as he rotated the prism through successive angles of 90°. He said that the light was *polarised;* for, favouring the corpuscular theory, he concluded that the corpuscles possessed *poles*, which gave rise to the observed effects. (The plane of reflection of the polarised light is called the

plane of polarisation.) Extending his investigation, he found that the light which is reflected from the surface of any transparent medium, at definite angles (called the angles of polarisation) which depend upon the nature of the medium, exactly resembles one of the beams which have passed through a doubly reflecting substance.

The reflected light has the same properties as the ordinary ray in Iceland spar has if the plane in which it is reflected is parallel to the optic axis of the spar; it manifests the properties of the extraordinary ray if its plane of reflection is perpendicular to the axis.

The supporters of the undulatory theory at first regarded the vibrations as taking place in the direction in which the waves travelled, but the phenomena of polarisation cannot be explained on this assumption. In particular, the conditions which are essential to the production of interference of polarised light (§ 247) necessitate the assumption that the vibrations take place perpendicularly to the direction of the ray.

Hooke, in 1672, had suggested that the vibrations occur in directions which are perpendicular to the ray; but the idea was never developed until its truth was inferred by Young and Fresnel, independently, not long after Malus had discovered that light was capable of undergoing polarisation by reflection.

A ray of light in which the vibrations of the ether all take place in one common direction, evidently possesses '*form*' or '*disposition*,' or '*sides*.' Think, for example, of a stretched cord placed between two smooth parallel planes which just touch it. Waves in which the vibrations are parallel to these planes can pass along the cord; perpendicular vibrations are incapable of existing. The waves possess '*sides*' which are in, and perpendicular to, the planes which confine the cord.

238. *Laws of Polarisation by Reflection and Refraction.*—(1) *Brewster's Law.* Brewster made an elaborate series of investigations on the angles of polarisation of various substances, with the object of connecting the phenomenon with other optical properties of the substances. He found that *the index of refraction is equal to the tangent of the angle of polarisation.*

From this law we can at once deduce the relation $\cos i = \sin r$, when i and r are the angles of incidence and refraction. Hence, *the refracted ray is perpendicular to the reflected ray.*

Since the refractive index varies with the wave-length, rays of different colours are polarised at different angles.

Jamin has found that the polarisation is not quite complete

except in some substances the refractive index of which is about 1·46.

(2) *Arago's Law.* The light which is refracted into a transparent medium is polarised to a greater or less extent. Arago found that *the quantity of polarised light in the refracted beam is equal to the quantity in the reflected beam, and the planes of polarisation of the two are at right angles to each other.*

Brewster's Law is applicable to reflection in the interior of a dense substance. Hence part of the unpolarised light in the refracted beam will undergo further polarisation when it is reflected at the second surface of the substance. If a sufficient number of parallel reflecting surfaces, such as those of a number of thin plates of glass placed one behind the other, be provided, the incident light may be divided into a reflected and a refracted beam, each of which is totally polarised in a plane at right angles to the plane of polarisation of the other. This arrangement constitutes a 'pile of plates.'

(3) *Malus's Law.* Light, which is incident at the polarising angle on a plane reflecting surface, is totally unaffected, as regards intensity, by a second reflection from a parallel plate of the same substance. But, if the plane of incidence upon the second plate be perpendicular to the plane of reflection from the first, the reflected beam will be totally extinguished. This subject was fully investigated by Malus, who found that *the intensity of the twice-reflected beam is proportional to the square of the cosine of the angle of inclination of the two planes of reflection.*

239. *Direction of Vibration in Polarised Light.*—The reflected ray, which (according to definition) is polarised in the plane of reflection, has symmetry with regard to that plane, since its intensity is totally unaltered by any number of reflections in that plane. It has also symmetry with regard to the normal to the plane of reflection, since it vanishes on reflection in any plane which passes through this normal.

We may therefore assume either that the direction of the vibrations in polarised light is perpendicular to the plane of reflection, or that it lies in the plane of reflection. Fresnel, in his theoretical investigations, made the former assumption; Maccullagh and Neumann adopted the latter.

The truth of the former is indicated by a number of considerations.

The vibrations of the ordinary ray in Iceland spar will be perpendicular to the optic axis provided that the vibrations of a ray polarised by reflection are perpendicular to the plane of polarisation;

and thus the uniform speed of that ray, in all directions, is readily accounted for. But, on the alternative assumption, the properties of the ordinary ray would be exceedingly difficult of explanation.

The ordinary and extraordinary rays produced by transmission through certain crystals, such as tourmaline, are coloured. When the two rays pursue nearly the same paths, identical colours are exhibited in each; and this occurs when the two rays traverse the substance nearly in the direction of the optic axis, so that their vibrations are nearly perpendicular to it. As the rays separate out from the axis, the colour of the ordinary ray remains constant, while that of the extraordinary changes greatly. Haidinger remarked that this favours the assumption that the vibrations of the ordinary ray are normal to the optic axis, and therefore take place along the normal to the plane of polarisation.

If a horizontal beam of polarised light, the vibrations of which are in lines inclined at an angle α to the vertical, falls perpendicularly on a diffraction grating, the lines of which are vertical, the direction of vibration in the diffracted beam will make with the vertical an angle β, which differs from α. Let a be the amplitude of the incident vibration. The resolved part of it parallel to the lines of the grating is $a \cos \alpha$; and the part at right angles to this is $a \sin \alpha$. If the diffracted beam makes an angle ϕ with the normal to the grating, the part of $a \sin \alpha$, which is perpendicular to the diffracted beam, is $a \sin \alpha \cos \phi$, and it is this part alone which is effective in the propagation of light. Hence the tangent of the angle which the new direction of vibration makes with the lines of the grating is $\tan \beta = a \sin \alpha \cos \phi / a \cos \alpha = \tan \alpha \cos \phi$. The angle β is therefore less than α. Consequently, if the plane of polarisation is perpendicular to the direction of vibration, the plane of polarisation of the diffracted beam will be more nearly perpendicular to the lines of the grating than that of the incident beam; and the reverse will happen if the direction of vibration lies in the plane of polarisation.

This result was deduced from theory by Stokes. He also tested it experimentally, and found that the result seemed to support Fresnel's assumption.

Another test, also due to Stokes, is based upon the nature of the polarisation of light which has undergone reflection from very small material particles.

Stokes remarks that no conclusion can be drawn so long as the particles are large compared with the wave-length of light, for then reflection occurs as it would from the surface of a large solid; but, when the particles are small compared with the wave-length, it

seems plain that the vibrations in the incident and the reflected rays cannot be at right angles to each other.

The small particles with which he experimented were obtained by highly diluting some tincture of turmeric with alcohol and adding water. A horizontal beam of sunlight fell upon the particles, and the light was found to be polarised in the plane of reflection. The smaller the particles were, the greater was the tendency to complete polarisation in the plane of reflection.

Since the 'sides' of the reflected ray are symmetrical with respect to the plane of polarisation, its vibrations must either be parallel to the incident ray or perpendicular to the plane of reflection, *i.e.*, of polarisation. We must therefore choose the latter alternative, since we cannot suppose that the directions of vibration in the incident and the reflected rays are at right angles to each other. (See, further, Chap. XXXIII.)

240. *Reflection and Refraction of Polarised Light.* — Young first determined the relations existing amongst the intensities of the incident, the reflected, and the refracted beams when light falls perpendicularly upon the bounding surface of two transparent media.

Fresnel, starting from certain assumptions, gave a complete investigation of these relations for all angles of incidence.

He assumed, first, the conservation of *vis viva* (or energy); second, continuity of displacement of the particles of the ether at either side of the bounding surface; third, proportionality of the density of the ether in a given medium to the square of the refractive index of that medium.

The third assumption implies that the rigidity of the ether (regarded as possessing properties analogous to those of an elastic solid) is the same in any two media. For the refractive indices are inversely as the speeds of propagation of light in the two media, and therefore the densities are inversely as the squares of the speeds. But (§ 168) the squares of the speeds are in direct proportion to the ratios of the rigidity to the density of each medium, from which it follows that the rigidity of the ether in each must be the same.

Let us suppose that light, polarised in a plane which makes an angle θ with the plane of incidence, is reflected from a transparent surface, and let a be the amplitude of its vibrations. The resolved parts of this parallel, and perpendicular, to the plane of incidence are $a \sin \theta$ and $a \cos \theta$, respectively. For shortness, let us denote these by p and q; and let p' and q' denote the similar portions of the amplitude of the reflected ray, while m and n represent the similar portions of the refracted ray.

The fact of conservation of energy is expressed (§ 161) by the equations

$$v\rho \cos i\,(p^2 - p'^2) = v'\rho' \cos r \,.\, m^2, \quad \ldots\ldots (1)$$
$$v\rho \cos i\,(q^2 - q'^2) = v'\rho' \cos r \,.\, n^2, \quad \ldots\ldots (2)$$

where ρ and ρ' represent the densities of the ether in the media. But, by Fresnel's third assumption, we have

$$\frac{\rho'}{\rho} = \frac{\sin^2 i}{\sin^2 r} = \frac{v^2}{v'^2},$$

whence (1) and (2) become respectively

$$p^2 - p'^2 = m^2 \tan i \cot r \quad \ldots\ldots\ldots\ldots (3)$$
$$q^2 - q'^2 = n^2 \tan i \cot r \quad \ldots\ldots\ldots\ldots (4).$$

The continuity of the displacement, in the case of vibration parallel to the surface, necessitates the condition

$$q + q' = n; \quad \ldots\ldots (5)$$

while, in the case of the vibrations which take place in the plane of reflection, it necessitates the condition

$$(p + p') \cos i = m \cos r \quad \ldots\ldots (6).$$

Combining (3) and (6), (4) and (5), we obtain

$$q' = -q \frac{\sin (i - r)}{\sin (i + r)} \quad \ldots\ldots\ldots\ldots\ldots (7),$$

$$n = \frac{2q \cos i \sin r}{\sin (i + r)} \quad \ldots\ldots\ldots\ldots\ldots (8),$$

$$p' = -p \frac{\tan (i - r)}{\tan (i + r)} \quad \ldots\ldots\ldots\ldots\ldots (9),$$

$$m = \frac{2p \cos i \sin r}{\sin (i + r) \cos (i - r)} \quad \ldots\ldots\ldots\ldots (10).$$

If $\theta = 0$, so that the incident light is polarised in the plane of incidence, and if the incidence is perpendicular, (7) shows us that the ratio of the intensity of the reflected, to that of the incident, light is

$$\frac{q'^2}{q^2} = \frac{\sin^2 (i - r)}{\sin^2 (i + r)} = \left(\frac{i - r}{i + r}\right)^2 = \left(\frac{\mu - 1}{\mu + 1}\right)^2.$$

The same equation shows that, when $i = 90°$, the whole of the incident light is reflected.

Again, by (8), when $\theta=0$, $i=0$, we see that the intensity of the refracted light bears to that of the incident the ratio

$$\frac{n^2}{q^2}=\left(\frac{2r}{i+r}\right)^2=\left(\frac{2}{\mu+1}\right)^2.$$

Similar expressions may be obtained from (9) and (10). All these results have been verified experimentally.

Let the plane of polarisation of the reflected light make an angle ϕ with the plane of incidence. From (7) and (9) we get

$$\tan\phi=\frac{p'}{q'}=\frac{p}{q}\frac{\cos(i+r)}{\cos(i-r)}=\tan\theta\,\frac{\cos(i+r)}{\cos(i-r)}.$$

Hence the plane of polarisation is rotated by reflection so as to more nearly coincide with the plane of incidence. The angle ϕ is equal to θ at perpendicular incidence; and it diminishes as i increases, until when $i+r=90°$—*i.e.*, at the polarising angle — it becomes zero. When i increases beyond this value, ϕ becomes negative; and, at grazing incidence, $\phi=-\theta$. So long as ϕ and θ have the same sign, the difference of phase of the two components of the reflected vibration is zero: at the polarising angle the difference changes suddenly to π.

If ψ be the angle which the plane of polarisation of the refracted light makes with the plane of incidence, (8) and (10) give

$$\tan\psi=\frac{m}{n}=\frac{p}{q}\sec(i-r)=\tan\theta\sec(i-r).$$

The rotation of both the planes of polarisation may be increased by successive repetitions of the same process. Since $\sec(i-r)=\sec(r-i)$, refraction *through* a parallel plate gives $\tan\psi_2=\tan\theta\sec^2(i-r)$.

Equations (5) and (6) above express the condition that there shall be continuity of displacement of the ether parallel to the bounding surface. No account has been taken of the displacement perpendicular to the surface. If we replace Fresnel's assumption of uniform rigidity by the condition of no normal discontinuity, we find $\rho=\rho'$. Hence Fresnel's third assumption is inconsistent with normal continuity of displacement.

Making the assumption that the ether is of uniform density in all media, Maccullagh and Neumann deduced expressions for the amplitudes of the components of the vibration of the reflected ray, in which the quantities on the right-hand sides of (7) and (9) are

simply interchanged. Hence, on this theory, we must assume that the vibrations are *in* the plane of polarisation. A similar interchange occurs in the expressions for the components of the vibration in the refracted ray; and, in addition, the magnitudes are altered. But this alteration occurs in such a way that the amplitude of vibration in the first medium, after refraction through a parallel plate, is identical on both theories. And, further, the rotations of the planes of polarisation are of the same magnitude and sense on the two theories. Therefore none of the phenomena with which we are now dealing are capable of furnishing a test between the assumptions of uniform density and uniform rigidity.

241. *Plane, Circular, and Elliptic Polarisation.*—In the special examples considered in last section, the difference of phase of the two rectangular components of the resultant vibration was 0 or π. But the resultant of two rectangular simple harmonic motions is, in general (§ 52), elliptic motion. Not only is this true of two rectangular components; it is true of any number of simple harmonic components in lines inclined at any angles to each other.

Hence, *if we can assume that the vibrations of a particle of the ether are simply harmonic when polarised light (such as we have hitherto considered) is passing*, we must conclude that the most general vibration of such a particle, when subject to various simultaneous disturbances, is elliptical. This assumption is justified by the fact that no phenomena of light, which are not due to simple superposition of displacements, are observed.

The resultant elliptic path is described continuously so long as the amplitudes, phases, and periods of the components remain constant. In this case the light is said to be *elliptically polarised.* As a particular case, when all the components can be compounded into two rectangular components, equal in amplitude and period, but differing in phase by $\pi/2$, the ellipse becomes a circle, and the light is *circularly polarised.*

Ordinary polarisation—for example, that produced by reflection—occurs when the components can be reduced to two which differ in phase by any multiple of π. This is usually termed *plane polarisation*, in order to distinguish it from the above forms.

242. *Nature of Common Light.*—Common light exhibits no trace of polarisation of any description. But this is known to be true also of plane polarised light if its plane of polarisation be made to rotate very rapidly—so rapidly that, in little more than one-tenth of a second, the directions of vibration have been distributed uniformly in all possible orientations perpendicular to the ray. Hence we may conclude that ordinary light consists of elliptically polarised

light, the magnitude, form, and position of the ellipse being in a constant state of rapid change.

But the phenomena of interference of light show that practically no change occurs in the course of some thousands of vibrations, for many thousands of interference bands can be counted when homogeneous light is used. On the other hand, since light travels at the rate of 186,000 miles per second, while the length of a wave is, on the average, about one forty-thousandth part of an inch, many millions of millions of vibrations must take place per second. But, again, as Stokes has pointed out, from the facts that every source of common light consists of a practically infinite number of points, and that the light emanating from each of these points is, in general, totally independent of that issuing from any other in respect of direction of vibration and also in respect of phase, we cannot expect anything else than an average effect in which there is no manifestation of 'sides.'

It follows that a beam of common light must necessarily be divided into two beams of equal intensity when it is transmitted through a doubly refracting substance.

From the formulæ (7) and (9) of § 240, we see that the total intensity of the reflected portion of a beam of unit intensity which falls on a transparent substance is

$$\frac{1}{2}\left(\frac{\sin^2(i-r)}{\sin^2(i+r)}+\frac{\tan^2(i-r)}{\tan^2(i+r)}\right),$$

since the two oppositely polarised parts into which the incident beam is supposed to be divided are of equal intensity. Now the second term in this expression is, in general, smaller than the first, and so, in the reflected beam, there is an excess of light polarised in the plane of incidence. The second term vanishes when $i+r=90°$, from which it follows that the reflected light is entirely polarised in the plane of incidence at the polarising angle.

Similarly, we can show that the refracted beam contains an excess of light polarised perpendicularly to the plane of incidence, that it is entirely polarised in the perpendicular plane at the polarising angle, and that its intensity is then equal to that of the reflected beam. At all angles of incidence there are equal amounts of polarised light in the two beams.

The known laws of the polarisation of common light by reflection and refraction are therefore consequences of the undulatory theory.

243. *Metallic Reflection.*—Malus observed that light is never completely polarised by reflection from the surface of metals, but

that the polarisation attained a maximum at a certain angle of incidence. He also observed that polarised light appeared to be completely depolarised by reflection from a metallic surface when its plane of polarisation was inclined at an angle of 45° to the plane of incidence.

Brewster verified, and extended, these results. He showed that the reflected portion of a beam of common light might be completely polarised by a sufficient number of reflections under like conditions—a result previously inferred by Biot. He found also that, when the incident ray is polarised in, or perpendicular to, the plane of incidence, the reflected ray is still polarised in the same plane; that, when the original polarisation is in any plane other than these, partial depolarisation seems to take place; and that the depolarisation is greatest at the angle of maximum polarisation. Further, a second reflection, in the same plane, and at the same angle, repolarises the light; and the new plane of polarisation lies on the opposite side of the plane of incidence and makes a different angle with it.

The 'depolarisation' above spoken of does not mean restoration to the condition of common light. The originally polarised light may be decomposed into two parts, one polarised in the plane of incidence, the other polarised in the perpendicular plane. The amplitudes of these parts may suffer change by reflection, which (§ 240) produces a rotation of the plane of polarisation. The phases may also be altered, and this will give rise to elliptic polarisation. Jamin's experiments on this subject show that Brewster's 'depolarisation' is really elliptical polarisation, and also show the nature of the variations of amplitude and phase.

The laws of change of amplitude are the same as those already found in § 240. The difference of phase increases from perpendicular incidence to grazing incidence by the total amount π, the phase of the ray polarised in the plane of incidence being accelerated with reference to that of the other. The change of phase is extremely slow, except in the immediate neighbourhood of the angle of maximum polarisation, between near limits on either side of which the total change occurs.

The difference of phase ought, according to Fresnel's theory (§ 240), to increase suddenly by π at the polarising angle.

Extending his observations to transparent bodies, Jamin found that the difference between them and metals is only a difference of degree.

In all cases elliptic polarisation is produced, and the maximum ellipticity occurs at the angle of maximum polarisation, which coincides very closely with the angle deduced from Brewster's law.

Jamin found that some transparent substances differ from metals with respect to the sign of the difference of phase which is produced by reflection. Substances whose refractive index is less than 1·46 *retard* the phase of the component which is polarised in the plane of incidence; substances which have a refractive index exceeding 1·46 resemble metals in accelerating the phase of this component; substances in which the refractive index is equal to 1·46 obey Fresnel's laws.

He also found that, in metals, the angle of maximum polarisation decreases as the wave-length of the light increases; from which we see that metals, if they obey Brewster's law, must refract light of long wave-length more than light of short wave-length. Recent experiments on refraction through thin metallic prisms seem to confirm this conclusion.

When light polarised in the plane of incidence is reflected from a metallic surface, the intensity of the reflected beam is a minimum at the angle of maximum polarisation. Maccullagh pointed out that transparent substances, the refractive index of which exceeds $2+\sqrt{3}$, possess (according to his theory) a minimum reflecting power at a definite angle of incidence.

244. *Double Refraction by Biaxal Crystals.* — Brewster discovered that most doubly refracting crystals possess *two* optic axes. In uniaxal crystals the axis is equally inclined to the three edges which meet at an obtuse-angled corner of the crystal. In biaxal crystals the lines which bisect the two angles contained by the axes, and the line at right angles to these two, have a definite relation to the crystalline form.

Fresnel has proved, theoretically and experimentally, that neither of the two rays in a biaxal crystal obeys the ordinary law of refraction. By means of certain assumptions, he investigated the problem of the propagation of waves of transverse vibration in a non-isotropic elastic medium. Crystalline substances are known to be non-isotropic, and presumably the property is impressed upon the ether which pervades them, and which is known to be hampered, as regards its free oscillation, by the presence of material particles.

The complete laws of double refraction may most readily be studied from the point of view of his theory, which we now proceed to give.

245. *Fresnel's Theory of Double Refraction.*—Fresnel undertook his investigation when the discovery of double refraction in biaxal crystals made it apparent that Huyghens' construction for the wave-surface was not applicable in all cases.

In a non-isotropic substance, the resultant force which opposes

the displacement of a particle does not in general act in the direction of the displacement. But Fresnel showed that there are three directions, at right angles to each other, in which the force, called into existence by the displacement, acts so as to move the particle directly back to its position of equilibrium. He showed that if, from any point in the interior of the substance, lines be drawn with lengths proportional to the square roots of the elastic forces which resist displacement in the directions in which the lines are taken, the extremities of these lines will lie on an ellipsoid (called the ellipsoid of elasticity). The three principal axes of this surface are in the directions in which the force tends to move the displaced particle directly back to its undisturbed position.

The speed of wave-propagation in an elastic medium is proportional (*cf.* § 168) to the square root of the elastic force (or distortional rigidity), and hence the radii of the ellipsoid are proportional to the speeds of propagation of waves when the vibrations are along the given radii.

Consider a plane wave passing through the medium. Fresnel proved, from the fact that the intensity of a beam of light which is compounded of two beams polarised at right angles to each other is independent of the phase of either component, that the vibrations must lie in the wave-front. If, therefore, we regard a central section of the ellipsoid of elasticity by the wave-front, we see that there are only two directions—those of the two axes of the section—in which a displacement will give rise to a reverse force acting in a plane which is normal to the wave and which passes through the line of displacement; for Fresnel showed that the force acts in the normal to that central section of the ellipsoid which is conjugate to the direction of the displacement. In general, the force will have a component perpendicular to the wave-front; but, according to assumption, this produces no effect in the way of wave-propagation.

Corresponding to any given plane wave-front, there are therefore only two directions of vibration such that the elastic force developed by the displacement has an effective component entirely in the direction of the displacement. But this condition is *essential* to the propagation of a permanent wave. Hence a plane wave, incident upon such a medium, is, in general, broken up into two waves, which are propagated in different directions with speeds which are proportional to the radii of the ellipsoid drawn in these directions.

[The following extract from Stokes' 'Lectures on Light' will aid in the formation of clear ideas on this point:

'Now we have not far to go to find a mechanical illustration of

such a mode of action. Imagine an elastic rod terminated at one end, and extending indefinitely in the other direction. Let the rod be rectangular in section, the sides of the rectangle being unequal, so that the rod is stiffer to resist flexure in one of its principal planes than the other. Let this rod be joined on to a cylindrical rod forming a continuation of it which extends indefinitely. Conceive the compound rod as capable of propagating small transverse disturbances, in which the axis of the rod suffers flexure. Imagine a small disturbance, suppose periodic, to be travelling in the cylindrical rod towards the junction. It will travel on without change of type, even though the flexure of the axis be not in one plane. But to find what disturbance it excites in the rectangular rod, we must resolve the disturbance in the cylindrical rod into its components in the principal planes of the rectangular rod, and consider them separately. Each will give rise in the rectangular rod to a disturbance in its own plane, but the two will travel along the rod with different velocities. This illustrates the sub-division of a beam of common light falling on a block of Iceland spar into two beams polarised in rectangular planes, which are propagated in the spar with different velocities. Again, suppose the original disturbance in the cylindrical rod confined to one plane. If this be either of the principal planes of the rectangular rod, the more slowly or the more quickly travelling kind of disturbance, as the case may be, will alone be excited in the latter; and if the plane of the original disturbance be any other, the components into which we must resolve it in order to find the disturbance excited in the rectangular rod will in general be of unequal intensity, their squares varying with the azimuth of the plane of the original disturbance in accordance with Malus's law. This illustrates the sub-division of a beam of polarised light incident on Iceland spar into two of unequal intensity polarised in rectangular planes, and their alternate disappearance at every quarter of a turn. We see with what perfect simplicity the theory of transverse vibrations falls in with the elementary facts of polarisation discovered by Huyghens, standing in marked contrast in this respect with the conjecture by which Huyghens himself attempted to account for double refraction.']

In any one direction in the interior of the substance two plane waves may be propagated, generally with different velocities; and the vibrations in these waves are necessarily at right angles to each other. But there are two directions in which the speed of propagation of a wave is independent of the direction of vibration. These directions are parallel to the normals to the two sets of circular sections of the ellipsoid of elasticity; for, all the radii of a circular

section being equal in length, the same force of restitution is called into play by a given displacement along any radius.

The resolution of an incident beam into two rectangularly polarised beams which obey Malus's law, and the existence of two directions in which a polarised beam is transmitted without modification, are therefore consequences of Fresnel's theory.

The planes of polarisation pass through the normal to the wave and through the major and minor axes of the section of the ellipsoid of elasticity by the wave. But the lines in which the circular sections cut the given plane section are equally inclined to the principal axes. Hence *the planes of polarisation bisect the dihedral angles which are contained by the planes which pass through the normal to the wave and the optic axes.*

The form of a wave which spreads out through the medium from any centre is formed by the following construction, which is due to Fresnel: Along the normals to any central plane section of an ellipsoid, similar to the ellipsoid of elasticity, measure from the centre lengths which are proportional to the principal axes of the section. The wave-surface is the locus of the points so found, and consists of two sheets. The directions of the refracted rays are found from this surface by Huyghens' construction. The tangent plane drawn to one sheet gives the direction of the one ray; that drawn to the other sheet gives the direction of the second ray.

The wave-surface possesses symmetry with respect to the three principal planes of the ellipsoid from which it is derived. Its traces upon these planes consist respectively of a circle and an ellipse. Let OA, OE, OF (Fig. 146) be the three principal axes. AB is a circle of radius OA equal to the mean principal axis of the ellipsoid; CD and EF are circles, the radii of which are respectively equal to the least and the greatest principal axes. BC is an ellipse, the principal axes of which are equal to the two least axes of the ellipsoid, and so on.

When the ellipsoid of elasticity is one of revolution, the wave-surface consists of two separate sheets—a sphere and an ellipsoid—which have one axis in common. Hence Fresnel's theory explains the double refraction of uniaxal crystals.

It is very desirable to note that Fresnel's results, although they are all seemingly in complete accordance with observed facts, are not rigorous deductions from his assumptions. Green and Neumann proved that a strict investigation, based upon these assumptions, will lead only approximately to Fresnel's Laws. Green then deduced these laws as rigorous consequences of an originally more general theory, the generality of which was subsequently limited by

imposing a very probable condition; but this theory made it necessary to suppose that the direction of vibration is in the plane of polarisation. He then showed that a still more general theory would, by means of suitable assumptions, lead to the same laws, and to the conclusion that the vibrations are perpendicular to the plane of polarisation. Other theories also give like results; for example, Maccullagh's theory, which is identical in its results with Green's first theory. On this point Stokes remarks that the principle of transverse vibrations is common to all these theories, while Fresnel's Laws are the simplest which can suit the phenomena; so that their mutual agreement can merely be regarded as a confirmation of that principle, while none of the special assumptions made as to the nature of the luminiferous medium in the interior of

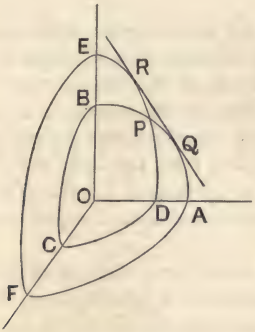


FIG. 146.

crystals can be regarded as being proved solely by the correctness of the results to which they lead.

246. *Conical Refraction.*—Sir W. R. Hamilton showed that the tangent plane QR (Fig. 146) touches the wave-surface at all points of a circle of contact, so that the point P is a 'conical point.' Four such points, one in each of the quadrants in the plane EOA, exist. They are, of course, the four points of intersection of the circle APB with the ellipse DPE.

The line OQ is perpendicular to the plane QR. But the perpendicular on the plane which touches the wave-surface represents the speed of propagation of the plane wave to which it is a normal; and the direction of the rays in the crystal are those of the lines joining O to the point of contact. Hence a plane wave, incident upon the crystal in such a direction that RQ is its front after refraction,

gives rise to a cone of rays which proceed in the directions of the lines joining O to the various points of the circle of contact. Hamilton's theoretical prediction of this phenomenon, which is known as *internal conical refraction*, was verified experimentally by Lloyd.

Both sheets of the surface are touched by the plane RQ, and hence there exists only one wave velocity in the direction OQ, which is therefore one of the optic axes. The other optic axis is the image of OQ in the plane EOF.

An infinite number of tangent planes may be drawn to the surface at the point P, and therefore a ray which proceeds through the crystal in the direction OP, gives rise on emergence to a thin conical sheet of rays. This *external conical refraction* was also predicted theoretically by Hamilton, and found, as the result of experiment, by Lloyd.

Since the radii of the wave-surface represents the velocities of the rays, we see that OP is a direction of single ray-velocity in the substance, for the two sheets of the wave-surface intersect at the point P. The other axis of single ray-velocity is the image of OP in EOF. The axes of single ray-velocity never deviate much from the axes of single wave-velocity, *i.e.*, from the optic axes.

247. *Interference of Polarised Light.*—The laws of interference of polarised light were investigated experimentally by Fresnel and Arago.

Two rays of light, which are polarised in perpendicular planes, do not give rise to phenomena of interference under circumstances in which rays of ordinary light would interfere.

The planes of polarisation of two such rays may, by suitable means, be made to coincide; but no interference will take place unless the two rays were originally parts of one polarised beam.

Rays polarised in the same plane will always interfere under the conditions in which rays of ordinary light would interfere.

When rays which have been polarised by double refraction produce interference, the phenomena exhibited are such as to necessitate the assumption that the phase of one component has been accelerated by the amount π relatively to that of the other. The reason for this is not far to seek.

Let OP (Fig. 147) represent the vibration in a ray which falls upon a crystal of Iceland spar the principal section of which is AA′. OP will be broken up into two components O*n* and O*m*, respectively along and perpendicular to the axis. Let each of these be again resolved in the directions *aa′* and *bb′*. The components along *aa′*

are in the same phase, but the components along bb' are necessarily in opposite phases.

248. *Colours of Crystalline Plates.*—Let us suppose that a beam of plane polarised white light is obtained by means of reflection, and let a second reflector be so placed as to extinguish the beam. If a thin crystalline plate be now interposed between the two reflectors in the path of the beam, intensely coloured light will in general be reflected from the second surface.

The light disappears whenever the principal section of the crystalline plate coincides with, or is perpendicular to, the plane of reflection from the first surface. If the plate be turned round, in its own plane, from this position, some reflection from the second surface will be evident, and the reflected light will vary in intensity,

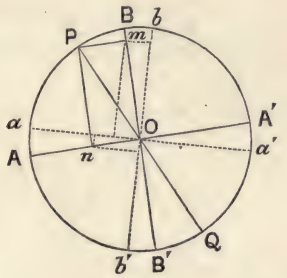


FIG. 147.

but not in colour, as the plate is rotated. (The same appearances, *colour excepted*, would, of course, be manifested even if the plate were thick.) On the other hand, if the plane of reflection at the second surface be varied, the crystal being fixed, the colour changes gradually into the tint which is complementary to the former. When any two successive positions of the second plane of incidence differ by 90° from each other, the reflected tints are complementary.

The colours depend upon the thickness of the crystalline plate, and vary with the thickness in the same way as the colours seen by reflection from a thin plate of air.

The explanation of the phenomenon is simple. The originally plane polarised light is divided in the crystal into two beams, which are oppositely polarised, and which traverse the plate with unequal speeds. The phase of the one beam is therefore retarded relatively

to that of the other, and so interference may take place if the vibrations in the two portions are again resolved in a common direction. The tint is due to the cutting out of some rays, and the intensifying of others, by interference.

The light which emerges from the plate is elliptically polarised, since it is compounded of two rectangularly polarised beams which differ from each other in phase. In particular, when the difference of phase is any odd multiple of a quarter of a period, the light is circularly polarised; and, when the difference is a multiple of half of a period, the light is plane polarised. If the difference is an even multiple of a half-period, the plane of polarisation coincides with the original plane; it makes an equal angle with the principal section, on the opposite side, when the difference is an odd multiple of a half-period.

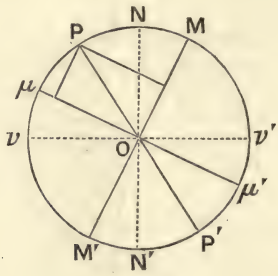


FIG. 148.

Let P O P′ (Fig. 148) represent the direction of vibration in the incident beam of light which falls upon the plane surface P O M of a doubly refracting plate, and let $\mu\ \mu'$ represent the principal section of the crystal so that the vibrations in the ordinary ray are in the direction O M. If the amplitude of vibration along O P be unity, the amplitude of the vibrations in the ordinary ray is $\cos\theta$, where $\theta = \text{P O M}$; and the amplitude of those in the extraordinary ray is $\sin\theta$. Let $\nu\ \nu'$ represent the principal plane of a second doubly refracting crystal through which pass both the beams into which the original beam is divided.

The vibrations along O M and O μ give rise to two sets of vibrations along O N, the amplitudes of which are $\cos\theta\cos(\theta-\phi)$ and $\sin\theta\sin(\theta-\phi)$ respectively, where $\phi = \text{P O N}$. These two vibrations are the components of the vibrations of the ordinary ray which emerges from the second crystal.

The vibrations along O M and O μ also give rise to two sets of vibrations along O ν, the amplitudes of which are $\cos\theta \sin(\theta-\phi)$ and $\sin\theta \cos(\theta-\phi)$ respectively. And these two vibrations are the components of the vibrations of the extraordinary ray which emerges from the second crystal.

The intensity of the ordinary beam is therefore $\cos^2\theta \cos^2(\theta-\phi)$

$$+\sin^2\theta \sin^2(\theta-\phi) + 2\cos\theta \sin\theta \cos(\theta-\phi) \sin(\theta-\phi) \cos 2\pi\frac{l}{\lambda},$$

where l is the effective difference of path of the two rays in the thin crystalline plate, and λ is the wave-length. Similarly, the intensity of the extraordinary beam is $\cos^2\theta \sin^2(\theta-\phi)+\sin^2\theta \cos^2(\theta-\phi)$

$$+2\cos\theta \sin\theta \cos(\theta-\phi) \sin(\theta-\phi) \cos 2\pi\frac{l}{\lambda}.$$

These two expressions easily reduce to

$$\cos^2\phi - \sin 2\theta \sin 2(\theta-\phi) \sin^2\pi\frac{l}{\lambda},$$

and

$$\sin^2\phi + \sin 2\theta \sin 2(\theta-\phi) \sin^2\pi\frac{l}{\lambda}.$$

When light of various wave lengths is used, the sum of all quantities of the form $\sin^2\pi\, l/\lambda$ must be taken.

From these expressions we can deduce all the observed effects.

The sum of the two intensities is unity, and therefore the colours of the ordinary and extraordinary beams are complementary.

The second term in each expression is the quantity upon which the coloration depends. Hence all colour vanishes when

$$\theta=0 \text{ or } \frac{\pi}{2},$$

that is, when the principal section of the crystalline plate is parallel or perpendicular to the original plane of polarisation. It also vanishes when

$$\theta-\phi=0 \text{ or } \frac{\pi}{2},$$

that is, when the principal sections of the thin plate and the second doubly refracting crystal are parallel or mutually perpendicular. The reason for this is that, in each of these four cases, one of the components of both the extraordinary and the ordinary beams necessarily vanishes: and the reason for the colours of the two beams being complementary is that the difference of the phases of the components of the extraordinary beam necessarily

differs from that of the components of the ordinary beam by the amount π.

The coloration is greatest when $\sin 2\theta \sin 2(\theta-\phi)$ is a maximum. As the maximum value is unity, this necessitates

$$\theta=\frac{\pi}{4},\ \phi=0 \text{ or } \frac{\pi}{2}:$$

that is, the principal section of the crystalline plate must make an angle of 45° with the original plane of polarisation, and the principal section of the second doubly refracting crystal must be parallel, or perpendicular, to that plane.

The colours change through regularly recurring cycles as the quantity l, and therefore as the thickness of the plate, increases by successive equal stages.

249. *Special Cases.*—Hitherto we have assumed that the incident beam of light is parallel. We shall now suppose that a diverging beam of polarised light traverses a uniaxal crystalline plate the parallel plane faces of which are perpendicular to the axis.

A ray O P (Fig. 149) which passes perpendicularly through the plate suffers no change. It will pass through, or be stopped by, a second plate which is cut parallel to its optic axis according as the principal

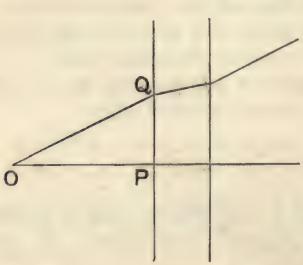


Fig. 149.

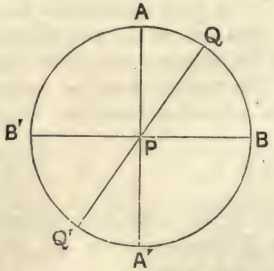


Fig. 150.

section of the second plate is parallel to, or perpendicular to, the original plane of polarisation. Any other ray will undergo change according to its inclination to the optic axis, and according to the angle which the plane passing through it and the axis makes with the plane of polarisation.

Let A B′ A′ B (Fig. 150) be perpendicular to the axis of the cone of rays, and let A A′ and B B′ represent respectively the original plane of polarisation and the perpendicular plane. All rays which

emerge from the crystal in these planes will be allowed to pass through, or will be stopped by, the second plate, according as its principal plane is parallel, or is perpendicular, to the original plane of polarisation. The field will therefore exhibit a light or a dark cross.

(*Haidinger's Brushes* are observed when polarised light is examined by the naked eye. The phenomenon consists of two yellowish-brown patches of light forming a brush the axis of which is parallel to the plane of polarisation; and two other bluish or purplish patches occur in the angles between the yellow patches. The appearance is due to a polarising structure which is most highly developed in dark eyes. It appears that the yellow spots of the eye are doubly refracting and absorb the extraordinary ray to a greater extent than the ordinary ray. Helmholtz finds that the effect only appears with blue light. The brushes soon disappear unless the plane of polarisation be changed at intervals.)

At any other point, such as Q, the vibration of the ray is resolved into its two components, polarised parallel and perpendicular to the plane through PQ. Thus the incident ray will be divided into two, which traverse the crystal with different speeds and so give rise to interference. The retardation of the phase of the one component relatively to that of the other is constant so long as the distance PQ is constant, and becomes greater and greater as PQ increases in length. Hence the field exhibits a series of alternately light and dark circles surrounding the point P.

The circles are brilliantly coloured if white light be used; and the

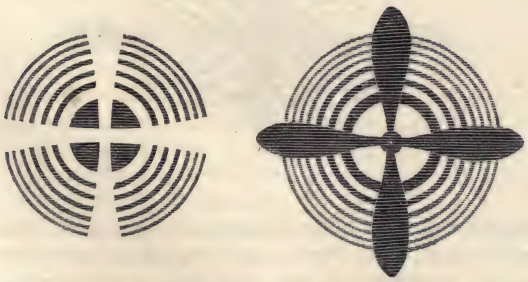


FIG. 151.

colours seen when the principal section of the second crystal occupies any definite position are exactly complementary to those which are seen when this plane is rotated through a right angle.

These effects are produced in comparatively thick crystals, since the difference between the speeds of the two rays, in directions not greatly different from the optic axis, is comparatively small.

The squares of the radii of successive circles are nearly proportional to the natural numbers. For it has been proved that the difference of the squares of the speeds of propagation of the two waves is proportional to the square of the sine of the angle which the ray within the crystal makes with the optic axis, and also that it is proportional to the thickness of the plate, and to the interval of retardation conjointly. Hence the retardation varies as the square of the sine of the angle between the ray and the axis. But this angle is very nearly equal to the angle QOP (Fig. 149), or to the distance QP.

Consider now a parallel plate cut from a biaxal crystal in a direc-

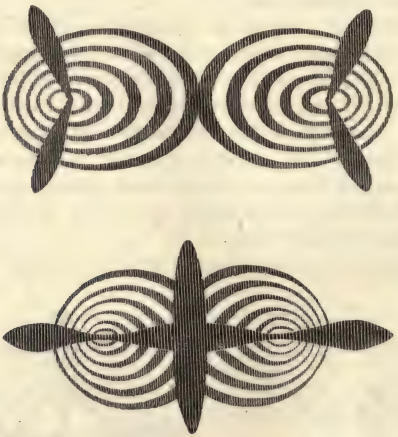


Fig. 152.

tion perpendicular to the line bisecting the optic axes. The interva of retardation is, in this case, proportional to the product of the sines of the angles which the wave normal makes with the axes. And these sines are approximately proportional to the distances of the point of emergence of the ray from the points in which lines drawn parallel to the axes, from the point of incidence on the first face of the crystal, meet the second face. Hence the locus of P is an oval of Cassini.

250. *Artificial Production of the Doubly Refracting Structure.*—Fresnel showed that glass and other singly refracting substances become doubly refracting when subjected to stress; and Brewster showed that unequal heating of the substance is sufficient to produce the requisite strain. Compression gives rise to a structure which resembles that of Iceland spar in producing an extraordinary ray the refractive index of which is less than that of the ordinary ray. Expansion produces an opposite effect.

The optic axis, in these cases, is fixed in *position* as well as in direction, unless the strain be homogeneous.

If an elliptic cylinder of glass be suddenly heated uniformly over its surface, a biaxal refracting structure is developed. This structure is also usually seen in unannealed glass.

Analogous changes may be produced in substances which naturally exhibit double refraction.

Clerk-Maxwell showed that a viscous liquid becomes doubly refractive while subjected to shearing stress.

251. *Rotatory Polarisation.*—Quartz is a doubly refracting substance in which the refractive index of the extraordinary ray is greater than that of the ordinary ray. The wave-surface consists of a sphere and a spheroid, but the spheroid lies entirely within the sphere. It follows from this that the speeds of propagation of the two rays along the optic axis are not the same. But, further, the vibrations in the two rays are not rectilinear. They are elliptical in general, and the ellipses are described in opposite directions in the two rays. When the rays are propagated along the axis, the vibrations become circular.

Now the resultant of two uniform motions, of equal period, in opposite directions in the same circle, is rectilinear motion. For, if A, A′ (Fig. 153) represent simultaneous positions of the oppositely moving points, the resolved parts of their motion perpendicular to the line PQ, which is drawn from the centre of the circle through the middle point of the arc joining A, A′, destroy each other. Hence the resultant is simple harmonic motion in the line PQ.

But if A′ be retarded relatively to A, the line PQ will take a new position, bisecting the arc between A and the new position of A′. And, if A′ is continuously retarded, PQ will revolve continuously round in the direction AA′. On the other hand, if A be retarded relative to A′, PQ will revolve in the direction from A′ to A.

The two circularly polarised rays in quartz therefore produce plane polarised light on emergence; but, because of the retardation of one ray relatively to the other, the plane of polarisation has been

rotated through an amount which is proportional to the thickness of the quartz.

The rotation is right-handed in some specimens of quartz; left-handed in others. Amethyst consists of alternate layers of right-handed and left-handed quartz.

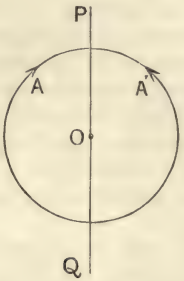


FIG. 153.

The amount of rotation is dependent upon the wave-length. Biot, and subsequently Broch, proved that it is approximately inversely proportional to the square of the wave-length. The first three terms of the formula

$$\rho = \frac{a}{\lambda^2} + \frac{b}{\lambda^4} + \frac{c}{\lambda^6} + \ldots\ldots$$

furnish a much better approximation, ρ being the rotation, and a, b, c being constants.

The spectrum produced by plane polarised sunlight which has passed through a plate of quartz is indistinguishable from that which is produced by ordinary sunlight. But a profound modification takes place if, previous to its passage through the refracting prism, the light be passed through an apparatus arranged to polarise light in a plane at right angles to the original plane of polarisation. Were the quartz plate taken away, no light could pass under these circumstances. The effect of the quartz is to restore the light, with the exception of such rays as have had their plane of polarisation rotated through a multiple of two right angles. Consequently, the spectrum is crossed by dark bands. By rotating the second polarising apparatus any one of the dark bands may be caused to occupy any desired position in the spectrum, and so the total rotation for any particular kind of light (and, consequently, the wave-length of that light) may be measured with extreme accuracy.

The dark bands are not sharply marked, for rays near those which are totally extinguished are necessarily partially extinguished. And, since portions of the light are cut out, the finally emergent light is coloured. The coloration disappears when the length of the quartz is so great that the portions which are cut out are distributed with practical uniformity throughout the spectrum.

Many liquids, solutions, and even vapours, possess this rotatory power—the rotation being in some cases left-handed, in others right-handed. As a rule, the rotation produced by a given thickness of a liquid is much smaller than that produced by the same thickness of quartz.

Neither dilution by an inactive substance nor vaporisation alters the rotatory power of a liquid, except in degree; on the other hand, Herschel showed that quartz is inactive when in solution; and Brewster showed that it is inactive when fused. Hence it is inferred that, while the rotation in quartz depends upon the crystalline structure, the rotation in liquids and vapours is a molecular phenomenon.

All rotatory polariscopes, or saccharimeters, depend essentially upon the above principles. In some of them, equal intensity of two beams of light is used as a test; in others, the test is furnished by the equality of tint of two beams. A spectroscopic test, such as that which has been described above, is by far the most delicate.

If the light which has passed through quartz be reflected back, so as to retraverse it in the opposite direction, the rotation will be undone. Rotation of the plane of polarisation may be produced in a magnetic field; but, in this case, reversal of the path of the light will not (Chap. XXXII.) be accompanied by an undoing of the rotation.

252. *Polarising Prisms.*—Many methods are used for the production of polarised light.

Plane polarisation by reflection has been already described.

Any doubly refracting crystal, of course, furnishes us with the means of producing plane polarised light, provided that we can separate the two beams.

Some substances, such as tourmaline, strongly absorb one of the two rays into which the incident beam is divided, and so furnish, when of sufficient thickness, a ready means of obtaining plane polarised light.

A block of Iceland spar separates the rays in proportion to its length, but it is extremely difficult to obtain large blocks which are free from internal flaws. In *Nicol's prism*, therefore, one of the two rays is got rid of by total reflection. A long block of the spar is divided into two parts by a plane which is perpendicular to its

principal section. The two portions are then cemented together, in their original position, by means of Canada balsam, the refractive index of which is intermediate in magnitude between the indices of the spar for the extraordinary and the ordinary rays. Consequently, when the inclination of the dividing plane to the path of the ray is sufficiently great, the ordinary ray suffers total reflection at the surface of the film of balsam, and the extraordinary ray alone is transmitted.

Foucault's prism is essentially similar, but the film of balsam is replaced by a film of air. A shorter block of the spar suffices in this method, but considerable loss of light takes place by reflection at the surface of the film.

Considerable separation of the rays may be obtained by the use of a prism of Iceland spar, or of quartz, which is achromatised by means of a prism of glass. In *Rochon's prism* the edge, and one of the faces, of the prism are perpendicular to the optic axis. The rays therefore pass through the prism without separation until they emerge at the opposite face. The prism is achromatised by means of a second prism, of the same substance, the refracting edge of which is parallel to the axis. The ordinary ray proceeds through the second prism without alteration of direction, and is therefore uncoloured; but the extraordinary ray is considerably refracted, and is coloured at its edges, since the refraction depends upon the wave-length.

Greater angular separation of the rays may be obtained by *Wollaston's prism*. In this arrangement the refracting edge of the first prism is perpendicular to the optic axis, but the face upon which the light falls perpendicularly is parallel to the axis. In all other respects the arrangement is the same as in Rochon's prism. Both rays are coloured at their edges, for both are deviated from their original direction.

Methods of producing elliptically, or circularly, polarised light have been already indicated. They all depend upon the introduction of a difference of phase between the two rectangularly polarised components of a beam of light. This difference may be produced by transmission through a doubly refracting plate. When the plate is of such thickness as to produce a difference of phase of a quarter-period, the light is circularly polarised provided that the two beams are of equal intensity. Such a plate is termed a quarter-wave plate. The difference of phase may also be produced by reflection or refraction—as in *Fresnel's rhomb*. This consists of a parallelepiped of St. Gobain glass the faces of which are so inclined that a ray enters and emerges perpendicularly at opposite faces after two

internal reflections at an angle of 54° 37′. Each reflection produces a difference of phase of 45°, and therefore the light emerges circularly polarised if it were originally polarised in a plane inclined at 45° to the plane of internal reflection.

Conversely, Fresnel's rhomb shows the existence of circularly polarised light by changing it into plane polarised light. This will also occur in the case of elliptically polarised light when either axis of the ellipse is inclined at 45° to the plane of internal reflection. But the two cases may be distinguished by means of a Nicol's prism; for rotation of the Nicol produces no alteration of intensity if the beam which is passing through it is circularly polarised, but it does produce variation of intensity if the beam is elliptically polarised. Though similar variation of intensity takes place if the incident beam is partially plane polarised, a quarter-wave plate, or a Fresnel's rhomb, enables us clearly to distinguish between these two cases.

CHAPTER XX.

THE NATURE OF HEAT.

253. *Radiant Heat. Its Identity with Light.*—We are accustomed to speak of the heat which we receive from the sun just as we speak of the light which we receive from it. And so the term 'radiant heat,' as applied to the heat which comes to us from distant bodies, apparently without the intervention of ordinary matter, has come into scientific use.

The non-intervention of ordinary matter in the process of radiation by which heat, like light, passes from one body to another at a distance from it, can readily be proved. There is no sufficient amount of ordinary matter in interstellar or interplanetary space to account for the transference: and a hot body cools in *vacuo* almost as readily as when surrounded by air—indeed, in certain cases, a hot body will cool less rapidly when surrounded by a material medium than it will otherwise.

Luminous bodies (those which exhibit fluorescence or phosphorescence being alone excepted) radiate heat as well as light, and the hotter they are the more luminous do they appear. We might, therefore, naturally conclude that the difference between light and radiant heat is not a difference of kind, but only a difference of degree. This inference is fully borne out by many facts.

One point in which they completely resemble each other is rectilinear propagation in free space or in a homogeneous medium. The heat-shadow which any obstacle casts is identical with the shadow which it produces with regard to light proceeding from the same source. And, since the path of light is rectilinear, this proves that radiant heat also moves in straight lines.

Heat, like light, is not propagated instantaneously; and an even more fundamental point of resemblance than the above appears in the fact that their speeds in *vacuo* are identical. This is made evident by the simultaneous disappearance and re-appearance of the light and heat when a total eclipse of the sun takes place.

The laws of reflection of the two are identical, for the focus of a mirror for heat-rays is the same as its focus for light-rays. A thermo-electric pile, placed at the focus of a reflecting telescope, can make evident the heat radiated from a star.

Their laws of refraction are also the same; although, at first sight, a difference appears because the focus of a lens for heat is farther off from the lens than its focus for light. But this only strengthens the analogy, for, unless the lens be achromatised, the focus for red rays is farther off than the focus for blue rays.

Both are governed by the same laws of interference and of polarisation. The phenomena of interference *prove* the existence of periodicity of motion, and show that the vibrations are transverse, as in the case of light. And the usual methods based upon interference, diffraction, and refraction, enable us to measure the wave-length, which is found to be greater than that of luminous rays.

We therefore conclude that light and radiant heat are one and the same thing; that the latter differs from the former only as red light differs from blue light; and that it is not evident to the sense of sight because the eye is so constituted that it cannot respond to the slower vibrations. We already know that some eyes can perceive rays at the red end of the spectrum (and also at the blue end) which are totally invisible to other eyes.

And colours—the word being used by analogy—appear in heat just as they do in light. Rock-salt is very transparent to heat rays just as glass is transparent to light; but, on the other hand, ordinary glass is very opaque to heat rays, *i.e.*, it absorbs them to a great extent. It acts to heat just as coloured glass acts to light; and many other substances act similarly. The law of absorption, with varying thickness of the medium, is the same as that which holds in the case of light (§ 205).

254. *Heat in Material Bodies. Hypothesis of Molecular Vortices.*—Since hot bodies give out radiation, and since the propagation of radiation involves motion of the particles of an inert medium, we might infer that the particles of a hot body must be in rapid motion, and that the communication of heat from one body to another depends upon the intercommunication of motion.

It is scarcely a century since Rumford and Davy arrived at this result upon experimental grounds.

Previous to their investigations heat was supposed to be a form of matter which was occluded in the interior of substances, but which was looked on as imponderable since it did not add to the weight of a body. This imponderable substance was termed *Caloric*.

According to the caloric hypothesis, a body was hotter or colder

in virtue of its having absorbed a greater or a smaller quantity of caloric; and when, by any means, the capacity of a body for caloric was diminished, it gave out heat (or rather caloric).

In the year 1798 Rumford was engaged in the boring of cannon, and *observed* (what had often previously been *noticed*) that there was a rise of temperature in the process of the reduction of the solid metal to the state of filings. But, according to the caloric hypothesis, the rise of temperature implies a diminution of the capacity of the substance for caloric; and, conversely, an increase of the capacity of a body for caloric would be accompanied by a fall of temperature unless additional caloric were supplied. Hence, in Rumford's experiments, the rise of temperature of the filings signifies, on this hypothesis, a diminution of the capacity for caloric.

Rumford, therefore, sought to determine by experiment whether or not the filings had less capacity for heat than the solid metal had. He heated equal weights of the solid metal and of the filings to the same high temperature, and dropped them into equal quantities of water at the same low temperature. He found that the same changes of temperature were produced in both cases, and concluded that the capacity of the substance for caloric had not been reduced when the body was broken up into smaller portions.

He was not aware that there was a distinct difference between the physical states of the two specimens of the substance which might have caused his experiment to indicate a wrong result. For the filings might have been so strained as to contain a considerable amount of latent heat which would only appear on their complete recovery from the state of strain. His conclusion was nevertheless correct, for it is entirely supported by experiments based upon accurate principles; and his observations therefore prove that the caloric hypothesis is incorrect.

But Rumford did not stop at this stage. He observed that the quantity of heat which was developed in the process of boring was independent of the amount of metal which was abraded—that a blunt borer and a sharp borer, though producing very different amounts of filings, caused the same development of heat if the same amount of work were spent in driving them. And, further, there seemed to be practically no limit to the amount of heat which might be produced. He therefore reasoned that heat could not be a material substance, and stated that he could scarcely conceive of anything but *motion* which could be excited and communicated in the manner observed.

Almost at the same time (in 1799) Davy was experimenting in precisely the same direction. He showed that two pieces of ice

might be melted simply by rubbing them together. Now heat is required to produce this change of state, and so, on the caloric hypothesis, the capacity of water for heat must be less than the capacity of ice for heat. But it is well known that the exact reverse is true.

The necessary heat might have been furnished by surrounding bodies, and therefore Davy enclosed the two pieces of ice by other portions of ice and placed them in the exhausted receiver of an air-pump. Under these conditions heat could only reach them by first melting the surrounding ice.

Davy was entitled to conclude, from the result of his experiments, that heat was not a form of matter, but, at the time, he merely said, 'Friction, consequently, does not diminish the capacities of bodies for heat.'

In 1812, when again discussing the point, he spoke of heat as 'a peculiar motion, probably a vibration, of the corpuscles of bodies tending to separate them,' and said that 'The immediate cause of the phenomenon of heat, then, is motion, and the laws of its communication are precisely the same as the laws of communication of motion.'

The second statement is rigidly correct; the first—that heat is motion—is only true if properly interpreted. Heat, since it is not matter, must be *energy;* and so the true meaning of Davy's statement is that heat consists in the energy of motion of the particles of a material body. But the word 'energy' was not introduced into science at that time.

Davy illustrated this heat-motion, which he termed 'repulsive motion,' by the analogy of the orbital motion of the planets. If the speed of motion of any planet were increased, the orbit would become larger just as if a repulsive force had acted.

In Davy's statement we have therefore the complete foundation of the whole kinetic theory of gases (Chap. XIII.), and, more generally, of the modern dynamical theory of heat.

Since heat is a form of energy we might infer the possibility of its existence in a potential form, and the use of the common term *latent* heat bears out the inference.

In working out a dynamical theory of heat Rankine advanced a hypothesis of 'molecular vortices.' He supposed that the motions which constitute heat in bodies are vortical motions in atmospheres around nuclei, and that radiation consists in the propagation of vibratory motions of the nuclei under their mutual forces. The energy of the vortices is the amount of heat which bodies possess; and the absolute temperature of any body is the quotient of this energy by a definite

constant for each substance. The elastic pressure, according to dynamical laws, must be directly proportional to the vortical energy, and must be inversely proportional to the volume which the vortices occupy, except in so far as the mutual nuclear forces which exist in all non-perfect gases modify this result. Latent heat is the equivalent of work done in varying the dimensions of the vortical orbits when the volumes and shapes of the spaces which they occupy are altered. Specific heat is the equivalent of work spent in varying the vortical energy.

CHAPTER XXI.

RADIATION AND ABSORPTION OF HEAT.

255. *Prevost's Theory of Exchanges.*—The fact that the laws of communication of heat are precisely those of the communication of motion leads to the conclusion that motion of the particles of a hot body is not confined to the surface alone; and this conclusion is confirmed by the greater radiation which takes place from a thick plate, than from a thin plate, of a transparent substance, when both plates are at the same temperature. It indicates also that the radiation from a hot body is dependent upon the state of that body alone, and is not influenced by the presence of any other body, except in so far as it may cause an alteration in the thermal state of the former.

This is the essence of the Theory of Exchanges which was advanced by Prevost of Geneva, under the title of a theory of 'movable equilibrium of temperature.'

According to Prevost, two bodies, which are of different temperatures, and are placed in an enclosure which is impervious to heat, will both radiate heat. The hotter body will radiate at a greater rate than the other, until, by absorption, the temperatures of the two are equalised. After this, each will still radiate, but at precisely the same rate; so that the heat which one loses by radiation is balanced by that which it gains by absorption. This is the condition which Prevost termed a condition of movable equilibrium (or, as we would now call it, of kinetic equilibrium) of temperature.

256. *Stewart's and Kirchhoff's Extension of Prevost's Theory.*—As in Chap. XVII., we define the *absorptive power* of a body, under given conditions, for any definite radiation, as the fraction of the whole incident radiation of that kind which it absorbs; and we also define the *emissivity* of a body, at a given temperature, for any given radiation, as the ratio of the quantity of that radiation which

it emits to the quantity of it which is emitted by a black body under the same conditions.

And Stewart's proof, as given in Chap. XVII., leads to the result that *the emissivity and the absorptive power of a body, at a given temperature, for any radiation, are equal.*

It is needless to enter into any discussion of special cases further than those which have already been given (§ 203). Suffice it to say that, whatever be the quality and the quantity of the radiation emitted by any body under given conditions, absorption must exactly balance emission, in respect both of quality and quantity, if the given conditions are to be maintained. Previously to Stewart's investigations, it was known from the experiments of Leslie, De la Provostaye, and Desains, that the radiating and the absorbing powers of any one body were proportional to each other; that is to say, it was known that a good radiator was a good absorber, and that a bad radiator was deficient in absorbing power.

257. *Laws of Radiation of Heat at Constant Temperature.*—Early in the history of the subject, it was known that the radiation from a body at a given temperature depended upon the nature of the surface of that body. (This furnishes an additional analogy between light and radiant heat.)

Leslie constructed a hollow metal cube, one side of which was polished, while another was rough. A third side was covered with lampblack, and the fourth was coated with white enamel. Although the surfaces of the two latter sides were so very different, Leslie found that both radiated about equally well when the cube was filled with hot water. The radiation from the bright metallic surface was much smaller than that from any of the others, and the radiation from the rough metallic surface was considerably less than that from the enamelled and the blackened surfaces.

It has already been proved (§ 205) that the amount of any given radiation, which is emitted from a sufficient thickness of a substance of given radiating power, is equal to that which is emitted from a black body at the same temperature. It was shown that the amount which is transmitted through a plate of the substance, n units in thickness, is $R(1-\rho)^n$, where R is the total incident radiation and ρ is the absorption co-efficient. Hence the amount which is stopped by the plate is

$$R(1-(1-\rho)^n);$$

and therefore, by definition, the absorptive power—and, consequently, the emissivity—is

$$1-(1-\rho)^n.$$

As the temperature of a body rises, radiations of shorter and shorter wave-length are emitted, and the energy in each previously existing kind is increased.

258. *Heat Spectra.*—If the radiation from a luminous body be passed through a slit and a prism in the usual manner, a spectrum is obtained which enables us at once to discover the nature of the radiation. And we may measure the amount of energy in any given portion of the spectrum by allowing that radiation to fall upon a medium which entirely absorbs it, and is consequently raised in temperature to a measurable extent. But, in any such experiment, it is necessary first to make certain that absorption does not take place, to any appreciable extent, in the substance of the prism.

We might also indirectly analyse the radiation emitted by a given body by means of determinations of the absorption which the body exercises upon radiations of different wave-lengths; but such measurements are of little value unless the substance used is of definite chemical composition and physical structure. Some gases (*e.g.*, olefiant gas) exert powerful absorption on the heat-rays; others exhibit very little. The absorption produced by water vapour seems to be largely due to the dust nuclei (§ 277).

These remarks apply to the invisible portions of the spectrum also, whether these consist of rays of higher refrangibility, or of rays of lower refrangibility, than those which form the visible parts.

It is found that the invisible portions of the spectrum possess characteristics which are precisely analogous to those which appear in the visible parts.

We cannot, by means of the thermopile (§ 328), or of the bolometer (§ 343), which are the two most suitable instruments for the present purpose, determine the energy of the radiation of one definite wave-length. All that can be done is to measure the energy of the total radiation which is contained between rays of known wave-length, for the face of the thermopile and the metal strip of the bolometer necessarily possess considerable breadth.

In the case of a refraction spectrum, besides the difficulty regarding absorption by the substance of the prism, there is the difficulty of the crowding together of the rays towards the less refrangible end of the spectrum according to an unknown law.

If a diffraction spectrum be used, absorption may take place in the substance of the grating if it be made of glass; and, if it be a metallic grating, great uncertainty exists with regard to the nature of its action upon the invisible rays. And, although the dispersion be (§ 233) practically proportional to the wave-length, so that equal

breadths of the spectrum correspond to equal differences of wave-length, we must remember that the measuring instruments ought to determine the energy contained in portions of the spectrum which are bounded by rays the wave-lengths of which are in a constant ratio. Further, since we are dealing with radiations of *all* wave-lengths, we see that, at any one part of a diffraction spectrum, an infinite number of different radiations are superposed, because of the existence of an infinite number of spectra of different orders.

The difficulties of the problem have been largely overcome by Langley, to whom is chiefly due our recent great increase of knowledge regarding radiations of large wave-length. He has detected, by means of the bolometer, traces of heat in portions of the solar spectrum corresponding to wave-lengths about twenty-four times greater than those of the least refrangible part of the visible spectrum.

He has shown that the wave-length at which the maximum of energy in the spectrum exists diminishes as the temperature is raised. This result has been deduced from theoretical considerations by Michelson; and the numerical deductions from the theory accord very well with Langley's observations. The energy at a given part of the spectrum dies away in amount very rapidly when we pass from the maximum in the direction of decreasing wave-length; it dies away much more slowly as we pass in the opposite direction along the spectrum.

Cauchy's formula (§ 209) connecting the refractive index of a substance for a given radiation with the wave-length of that radiation does not agree with Langley's observations on the heat-rays. Briot's formula agrees better, but it ultimately differs from them in the opposite direction to that in which Cauchy's differs from them.

259. *Radiation at Different Temperatures.*—Hitherto we have assumed that the radiating bodies with which we have dealt have been kept at constant temperature so that their radiating powers remained unaltered. We must now consider the relation between emissivity and temperature.

Let us suppose that the hot body is placed inside an enclosure which is kept at a constant temperature t. Let $t+\theta$ be the temperature of the hot body, and let no heat pass from it except by radiation.

If $f(t)$ represent the rate of loss of heat from the hot body at temperature t, the rate of loss of heat under the assumed conditions will be

$$f(t+\theta) - f(t);$$

and, since we have no adequate knowledge of the mechanism of

emission, the form of this expression must be determined from experiment.

According to Newton, the rate of loss of heat is proportional to the excess of the temperature of the hot body over that of its surroundings; or, in symbols,

$$f(t+\theta)=f(t)+a\,\theta.$$

But, since we must regard the rate of emission as independent of the surrounding bodies, this is equivalent to

$$f(t)=a\,t+b,$$

where a and b are constants, and t is *any* temperature.

The above law is not even roughly accurate unless the differences of temperature are small, and it becomes less and less applicable the greater the differences are.

Dulong and Petit made an elaborate series of experiments with the view of discovering a more correct law. They found that, when the temperature difference was kept constant, the rate of loss increased in geometrical progression as the temperature of the surroundings of the body increased in arithmetical progression; and the ratio of the geometrical series is independent of the temperature excess—so long, at least, as the excess is not greater than 200° C.

When the temperature excess vanishes, the loss of heat is zero. Dulong and Petit therefore expressed the law in the form

$$f(t+\theta)-f(t)=a\,a^t(a^\theta-1),$$

which agrees very closely with the result of their observations when t varies from 0° up to 80° C., and θ does not exceed 200° C. Since this formula may be written in the form

$$f(t+\theta)-f(t)=a\,a^{t+\theta}-a\,a^t,$$

we see that the absolute rate of radiation, independently of the surroundings, is

$$f(t)=a\,a^t+b.$$

The constant a depends on the nature of the radiating surface, but the constant a is practically an absolute constant, and is equal to 1·0077. The law of Dulong and Petit, therefore, asserts that, when the temperature excess is constant, the rate of loss of heat is proportional to $(1{\cdot}0077)^t$, where t is the absolute temperature of the

bodies to which the heat is radiated. It asserts also that the rate of loss is proportional to $(1{\cdot}0077)^{\theta} - 1$ when t is constant.

The equation

$$[f(t_1) - f(t_3)] = [f(t_1) - f(t_2)] + [f(t_2) - f(t_3)]$$

or

$${}_1r_3 = {}_1r_2 + {}_2r_3,$$

where r represents rate of loss, might not have been borne out by experiment. The fact that it is so borne out furnishes, as Balfour Stewart pointed out, an independent proof of the truth of Prevost's Theory of Exchanges, of which it is a necessary consequence.

We may write the expression $a^{\theta} - 1$ in the form $(1 + p)^{\theta} - 1$, which, by expansion, becomes

$$p\,\theta\left(1 + \frac{\theta - 1}{2}p + \ldots\right).$$

But, by Newton's law of cooling, the rate of loss should be proportional to θ; and hence, substituting the values $\theta = 11$, $p = 0{\cdot}0077$, we see that Newton's value is fully 4 per cent. too small when the temperature excess is 11°.

More recent experiments by De la Provostaye and Desains verified the accuracy of the law of Dulong and Petit within the limits already assigned.

According to Hopkins the radiation per square foot per minute, from glass at 100° C. to an enclosure at 0° C., is 0·176 heat units, the unit being the amount of heat which is required to raise the temperature of one pound of water from 0° C. to 1° C. Under the same conditions the radiation from unpolished limestone is 0·236 units; and that from polished limestone cut from the same block is 0·168 units.

Dulong and Petit's law seems to be applicable only to the total radiation from a body, and not to each definite radiation of which the whole is composed. The rate of emission of particular radiations from a black body seems to increase rapidly at first, and then more slowly, as the temperature of the black body is raised.

260. *Solar Radiation.*—Pouillet was the first to make fairly accurate measurements of the amount of radiation which is received by the earth from the sun in a given time. For this purpose he invented the *Pyrheliometer*.

This instrument consists of a flat cylindrical metallic vessel, the surface of which, with the exception of one end which is covered with lamp-black, is highly polished. The bulb of a thermometer is

inserted in the cylinder, and its stem lies along the axis. The cylinder is filled with water or mercury, and its blackened face is directed towards the sun. In order that this may be done with accuracy, a metal disc, the diameter of which is exactly equal to that of the cylinder, is fixed on the end of the axis of the instrument remote from the cylinder. The shadows of the disc and the cylinder will not coincide unless the face of the latter be accurately turned towards the sun. The area of the blackened face, and the amount of heat necessary to raise the temperature of the cylinder and its contents to a given extent, are exactly determined.

If the temperature of the instrument is the same as that of the air, and if its blackened face, carefully shaded from the sun, be turned towards the sky, radiation will take place, and the temperature will fall $\theta°$ (say) in t units of time.

Now let the apparatus be turned towards the sun for t units of time; after which let it be turned towards the sky as before, during an equal time, and let θ' be the fall of temperature.

It is concluded that the deficiency introduced into the total rise of temperature which took place when the instrument was exposed to the sun, is

$$\frac{\theta + \theta'}{2};$$

for, during this exposure, the cylinder was steadily rising in temperature because of the heat which it absorbed from the sun, and was at the same time steadily falling in temperature because of its own radiation.

Hence the full rise of temperature, had no radiation taken place from the cylinder, would have been

$$\Theta + \frac{\theta + \theta'}{2},$$

where Θ is the rise of temperature which was actually observed. And so, from the known constants of the instrument, the quantity of heat which was received, in a given time, by unit area at the earth's surface could be calculated.

Pouillet gave the expression

$$a\, e^{l}$$

for the rate of rise of temperature under the above conditions at different times of the day. The quantity l is the thickness of the earth's atmosphere which was traversed by the sun's rays. The constant e varies from day to day, but the constant a does not so vary. Pouillet concluded that the expression would be applicable

even if there were no atmosphere, in which case the constant a would represent the rate of rise of temperature; and he calculated that, in the event of no atmospheric absorption, the quantity of heat falling on a square centimetre of the earth's surface, in one minute, would raise the temperature of 1·76 grammes of water by 1° C.

From the average values of the constant e, Pouillet inferred that about one-half of the total incident solar radiation is absorbed in the earth's atmosphere. Sir W. Thomson concludes, from the above data, and those independently given by Herschel, that the rate of radiation from the sun's surface is about 7,000 horse-power per square foot, or thirty times the amount which is radiated, per square foot, from the furnace of a locomotive.

The *Actinometer* is another instrument used to determine the magnitude of solar radiation. It consists essentially of a metal enclosure (blackened internally and kept at constant temperature) at the centre of which the bulb of a thermometer is placed. By means of an opening in the enclosure, the sun's rays are allowed to fall upon the bulb for a given time, after which the bulb radiates to the enclosure. Various experimenters have used this apparatus in one or other of its forms. By its means Violle has found that the radiation from the sun, which would fall per minute on a square centimetre of the earth's surface did no absorption occur, would raise the temperature of 2·54 grammes of water by 1° C.

Langley's more recent measurements also indicate that Pouillet's estimate is too low. He finds that the quantity of unabsorbed heat falling on a square centimetre would raise the temperature of 1·81 grammes of water by 1° C. Did no absorption take place, this would become 2·8 grammes, and the amount of heat radiated from the sun per square foot of its surface would be fifty times greater than that radiated from a square foot of the surface of a locomotive.

From Langley's data we can calculate that the yearly radiation from the sun to the earth would, if spread uniformly over the surface, melt a uniform crust of ice fully 150 feet in thickness.

261. *Radiation from Moving Bodies.*—The following considerations regarding the motion of radiating bodies, first advanced by Balfour Stewart, are specially worthy of notice—apart from their intrinsic value—as an example of the useful employment of the principle of Conservation of Energy.

Let us suppose that ultimate equality of temperature has been arrived at inside the enclosure, impervious to heat, which was postulated in § 203; and let us further assume that one of the bodies in that enclosure suddenly commences to move about with a speed

which is comparable with that of light. A direct application of Döppler's principle shows us that any other body in the enclosure, towards which the motion may be directed, will receive energy from the moving body at a greater rate than will one which is so situated that the motion is directed from it.

It follows that relative motion of radiating bodies is inconsistent with ultimate equality of temperature. But persistent inequality of temperature would imply a perpetual source of energy; and we therefore conclude that the relative motion of radiating bodies must gradually cease.

CHAPTER XXII.

EFFECTS OF THE ABSORPTION OF HEAT: DILATATION AND ITS PRACTICAL APPLICATIONS.

262. *Temperature.*—Increase of temperature usually accompanies the application of heat to any substance. The fundamental distinction between *heat* and *temperature* is very clearly brought out by a consideration of the meaning of the words *hot* and *cold* as applied to different material substances. The bodies which are said to be hot or cold may really be at the same temperature. Thus a mass of iron and a mass of wood, though their temperatures be equal, *feel* very differently to the touch. If the temperature of the hand be higher than that of the two masses, the former feels cold, and the latter feels warm; while, if the temperature of the hand be lower than that of the bodies, these conditions are reversed. The reason s that the physical properties of the substances are such that, of the two, iron is the one which most rapidly abstracts heat from, or supplies heat to, the hand.

It is sufficient for our present purpose that we regard temperature as that condition which determines the flow of heat from one body to another. (See §§ 254, 150.)

If two bodies, which differ in temperature, be placed in contact with each other (or in thermal communication of any sort), heat passes from the body which is at the higher temperature to the body at the lower temperature. And two bodies, between which, *on the whole*, there is no transference of heat when thermal communication is established, are said to have *equal* temperatures.

It is an experimental fact that any two bodies, which have each the same temperature as a third body, are themselves at equal temperatures.

Various methods, which will be described later, are used for the determination of the difference of temperature which subsists between two bodies, or between a body in one given physical state and the same body when in another state.

We may premise, for present purposes, that the temperatures of ice-cold water, and of boiling water, respectively, are called 0 degrees and 100 degrees; that a given number of degrees, at any part of our scale, corresponds to a constant interval of temperature; and that we call the degrees *Centigrade* degrees—the letter C being used to discriminate them from the degrees of other scales.

263. *Dilatation of Solids.*—One of the most obvious effects of the application of heat to a substance is expansion. In a homogeneous isotropic solid, the expansion is equal in all directions. On the other hand, in a non-isotropic solid, the expansion is different in different directions; but, in this case, three rectangular directions (called the principal axes, *cf.* § 245), can always be found such that in one the expansion is a maximum; in another, it is a minimum; and, in the third, its value is intermediate between those along the other two — being, in fact, a maximum-minimum. When the expansions, along these directions, which accompany a given rise of temperature, are known, the expansion in any other direction can be found.

We have, first of all then, to consider the laws of linear dilatation of a solid. [The requisite measurements are readily made by direct micrometric methods, which give accurate determinations of the lengths of a bar at different known temperatures.]

1. It is found that *the increase of length of a given bar is proportional to the length of the bar at the initial temperature.*

2. *The alteration of length is proportional to the increase of temperature.*

These laws are symbolically represented by the equation

$$l_t = l_o(1+kt),$$

where l_t and l_o are respectively the lengths of the bar at the higher and lower of the two temperatures of which t is the difference, and k is a constant called the *co-efficient of linear dilatation.* This constant is obviously the increase of length, per unit rise of temperature, of a bar the original length of which is unity.

From the three equations

$$\begin{aligned} l_t &= l_o(1+k_1t), \\ b_t &= b_o(1+k_2t), \\ d_t &= d_o(1+k_3t), \end{aligned}$$

where the letters l, b, and d denote the length, breadth, and thickness of a rectangular parallelepiped of the substance, and k_1, k_2, k are respectively the co-efficients of dilatation measured in the direction of the length, breadth, and thickness, we can obtain an expression

for the cubical dilatation of the solid. Let us suppose that these equations refer to the principal axes; then

$$l_t b_t d_t = l_o b_o d_o (1+k_1 t)\,(1+k_2 t)\,(1+k_3 t),$$

or

$$\mathrm{V}_t = \mathrm{V}_o (1+k_1 t)\,(1+k_2 t)\,(1+k_3 t),$$

where the Vs represent volumes.

In all cases in nature, the quantities k_1, k_2, k_3 are so small that their squares and products may be neglected; hence, to a sufficient degree of approximation,

$$\begin{aligned}\mathrm{V}_t &= \mathrm{V}_o(1+(k_1+k_2+k_3)t)\\ &= \mathrm{V}_o(1+\mathrm{K}t),\end{aligned}$$

if K represents the *co-efficient of cubical dilatation*. This gives

$$\mathrm{K} = k_1+k_2+k_3;$$

and so the co-efficient of cubical dilatation (which is the fractional increase in bulk of unit volume per unit rise of temperature) is equal to the sum of the three principal co-efficients of dilatation. When the substance is isotropic, this latter equation becomes

$$\mathrm{K} = 3k;$$

which asserts that the co-efficient of cubical dilatation of a homogeneous isotropic solid is three times the co-efficient of linear dilatation.

If the edges of the rectangular parallelepiped be not in the directions of the three principal axes, the effect of the application of heat will be to change the inclination of the faces, so that the parallelepiped will cease to be rectangular. When the three edges are all originally equal, and the diagonals of one of the square faces of the cube are parallel to two of the principal axes, the application of heat changes the square face into a rhombus such that the tangent of half its obtuse angle is equal to $1+(k_1-k_2)t$, where t is the increase of temperature and the original length of the diagonals is assumed to be unity. This affords a ready means of determining the co-efficient of expansion along one of the two principal axes, provided that we know its value along the other; for the change of angle can be measured with extreme accuracy by optical methods.

Fizeau introduced a specially delicate method of measuring the change of length of a rod. Newton's rings (§ 221) are produced by the interference of light reflected at the surface of a film of air contained between two plates of glass, of which one at least is slightly curved; and the colour of any particular ring depends on the thickness of the film, which can be calculated when the radius

of the ring is known. If one of the two plates be fixed, while the other is attached to the expanding rod, the slightest change of length of the rod causes a diminution of the thickness of the film of air which can be easily calculated by means of the optical changes which are simultaneously produced.

He gives the following empirical formula connecting the co-efficient k with the temperature t expressed in Centigrade degrees:

$$k=a+a'(t-40).$$

The constants were determined by observations at three temperatures, viz., 10°, 45°, and 70°.

The cubical dilatation of a solid is usually determined directly by heating it in a liquid, of known expansibility, which is contained in a vessel made of a substance whose co-efficient of linear (and, therefore, also of cubical) dilatation has been found. Let V_o be the volume of the vessel at 0° C., and let K_1 be its co-efficient of dilatation. The volume at t° C. is

$$V_o(1+K_1t).$$

Similarly, the volume at t° C. of the liquid which filled the vessel at 0° C. is

$$V_o(1+K_2t).$$

If we now place a solid, the volume of which at 0° C. is v_o, and at t° C. is

$$v_o(1+K_3t),$$

the volume of liquid which overflows from the vessel at temperature t is

$$V_o(1+K_2t)-V_o(1+K_1t)+v_o(1+K_3t),$$

or

$$V_o(K_2-K_1)t+v_o(1+K_3t),$$

from which expression the value of K_3 may be found.

It is needless to enumerate the various practical applications of the dilatation of solid bodies when their temperature is raised, or of their contraction as the temperature falls. The well-known processes of shrinking the tires on wheels, and of drawing together the walls of a building when these have bulged outwards, will sufficiently serve as an indication of their nature.

One particular application—to the construction of a compensation-pendulum or balance-wheel of a watch—merits special notice.

The ordinary compensation-pendulum is constructed upon the principle that the *difference* between the lengths of two rods, of different expansibilities, will remain constant, however the temperature may be altered within allowable limits, provided that the lengths

of the rods be made in inverse proportion to the co-efficients of dilatation of the substances of which they are composed. This obviously affords a means of keeping constant the distance between the bob of a pendulum and its point of support.

If two equal bars, of different expansibilities, be soldered together throughout their length, a rise of temperature will cause the compound bar to bend in such a way that the less expansible bar is on the concave side; for this is the only way in which the tendency towards unequal expansions can be satisfied. This fact is made use of in the construction of compensated balance wheels. When the temperature of an uncompensated wheel increases, the expansion of spokes carries the rim of the wheel further out from the centre, and the consequent increase of moment of inertia produces an increase in the period of oscillation. Hence a watch, or chronometer, fitted with such a wheel will go too slow in warm weather and too fast in cold weather. But if the rim of the wheel be divided into a number of independent parts, each part being carried by a separate spoke, and if the rim be compound as above described, matters may be so arranged that the throwing-out of the weight from the centre because of the expansion of the spokes is counterbalanced by means of the inward bending of the segments of the rim.

It is worthy of note that one or two of the principal expansibilities of a solid may be negative, *i.e.*, the substance may contract in at least one direction when its temperature is raised. In such a case a series of directions may be found in the substance such that a change of temperature does not give rise to any alteration of length of the substance in any of these directions. [These directions may be found by imagining a sphere to be drawn in the unheated body and finding its intersection with the ellipsoidal surface into which it becomes distorted by the application of heat. All lines drawn in the body parallel to the lines joining the points of the curves of intersection to the centre of the sphere are unchanged in length.] Hence a rod cut from the substance in any such direction may be used as the rod of a compensated pendulum. Brewster pointed out that a rod of marble might be so used.

An interesting example of contraction of a solid when its temperature is raised is seen in the case of india-rubber under tension. The experiment may readily be performed by blowing steam through a hollow tube of the substance, which is fixed at its upper end, and is extended by means of a weight attached to the lower end.

The subjoined table contains Fizeau's determinations (see above) of the values of the constants in his formula for the co-efficients of linear dilatation of a few well-known substances. The constant a is,

of course, the value of the co-efficient at 40°. When more than one value is given these refer to the various principal dilatations.

	a.	a'.
Carbon (retort) ...	0·00000540	0·0000000144
Platinum	0·00000905	0·0000000106
Steel	0·00001095	0·0000000124
Iron (compressed) ...	0·00001188	0·0000000205
Copper (native) ...	0·00001678	0·0000000205
Silver	0·00001921	0·0000000147
Lead	0·00002924	0·0000000239
Quartz	0·00000781	0·0000000205
	0·00001419	0·0000000348
Iceland spar ...	0·00002621	0·0000000160
	0·00000540	0·0000000087
Arragonite	0·00003460	0·0000000337
	0·00001719	0·0000000368
	0·00001016	0·0000000064

An average value of the co-efficient of expansion of glass is 0·0000085.

264. *Dilatation of Liquids.*—In liquids it is merely the cubical dilatation with which we have to deal. This may readily be determined if we know the cubical dilatation, K′, of the substance of which the containing vessel is composed. Let K be the unknown co-efficient, and let an amount of the liquid, of weight W, fill the vessel (which must have a narrow neck) at a temperature, t, while the weight of the quantity which fills the vessel at 0° is W_0. Now the weight of the quantity which would have filled it, had the liquid been inexpansible, is $W_0(1 + K't)$; but, since the liquid is expansible, this amount is diminished in the ratio of unity to $1+Kt$. Hence we get

$$W(1+Kt) = W_0(1+K't),$$

which determines K.

A very simple method, by means of which the co-efficient of dilatation of a liquid may be found without any knowledge of that of the substance of which the containing vessel is composed, was devised by Dulong and Petit. In all essential particulars the apparatus consists of a double U-tube, *abcd* (Fig. 154). The portion *bc* is occupied by air, which separates the portions of the liquid contained in *ab* and *cd* while preserving continuity of pressure between them. Equal atmospheric pressure acts at the points *a* and *d*, and the air in *bc* is necessarily at uniform pressure throughout. Hence, when

equilibrium is maintained, the pressure per square inch due to the difference of level ab must be equal to the pressure per square inch due to the difference of level dc. Let us suppose that the limb cd

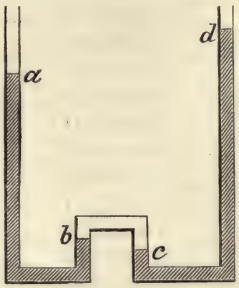


FIG. 154.

is raised to temperature t, while the limb ab is kept at 0°. The average expansibility throughout the range of temperature is

$$\frac{dc - ab}{ab \,.\, t}.$$

If dv be the increase of volume produced by a small rise of temperature dt, while v_0 is the total volume of the liquid at 0°, the quantity

$$\frac{dv}{v_0 dt}$$

is usually taken as the co-efficient of dilatation at the temperature t. The *true* co-efficient at temperature t is obviously given by the ratio

$$\frac{dv}{vdt},$$

where v is the volume of the liquid at that temperature.

The values, given by Regnault, of the ordinary co-efficient of dilatation of mercury at temperatures varying from 0° C. to 350° C. are well represented by the formula

$$K = 0{\cdot}0001791 + 0{\cdot}0000000504t.$$

Water presents a marked peculiarity as regards its change of volume with rise of temperature. Between its freezing-point, 0° C., and a temperature of almost exactly 4° C., its volume diminishes as the temperature increases. At all higher temperatures the volume

increases as the temperature is farther raised. Thus water is in a condition of *maximum density* at a temperature of about 4° C.

The existence of the temperature of maximum density is readily shown by means of *Hope's experiment*. The necessary apparatus consists of a cylindrical glass vessel the central portion of which is surrounded by a metal casing in which a freezing mixture may be placed. Two thermometers are inserted horizontally through the glass vessel, one near the top and the other near the bottom, so that their bulbs are at the axis of the cylinder. The vessel being filled with water, a freezing mixture is placed in the casing. Very soon the temperature marked by the lower thermometer begins to diminish, which shows that the cold water is of greater density than the warm water near the top of the vessel. This goes on until the lower thermometer registers a temperature of 4° C., at which it remains. Soon afterwards the temperature of the water at the top of the vessel begins to fall, which shows that the colder water is now ascending and must therefore be expanding; and this process goes on until the water at the top freezes at 0° C.

The maximum-density point is lowered by pressure to the extent of about 3° C. by a pressure of one ton's weight per square inch.

Kopp's determinations of the co-efficient of dilatation of water are fairly well represented between 0° C. and 20° C. by the formula

$$K = \frac{t-4}{72{,}000}.$$

The more recent experiments of Pierre, Hagen, and Mathiessen indicate that the denominator of this fraction is about 5½ per cent. too large; but Rossetti's results agree better with those of Kopp.

Different experimenters have determined the co-efficients of expansion of various liquids when under pressure sufficient to keep the liquid in equilibrium with its vapour at temperatures above the ordinary boiling-point. Drion gives the subjoined values of the co-efficient of dilatation of sulphurous acid:

Temp. C.	Co-efficient.	Temp. C.	Co-efficient.
0° ...	0·00173	70° ...	0·00318
10° ...	0·00188	90° ...	0·00415
30° ...	0·00219	110° ...	0·00592
50° ...	0·00259	130° ...	0·00957

From these results it appears that the co-efficient of dilatation of this liquid, at about 120° C., is double of that of air.

Hirn gives the volume of water at different temperatures as follows:

Temp. C.	Volume.	Temp. C.	Volume.
4° ...	1·00000	140° ...	1·07949
100° ...	1·04315	160° ...	1·10149
120° ...	1·05992	180° ...	1·12678

Consequently, at 180° C., the expansibility of water is nearly one half of that of air.

265. *Dilatation of Gases.*—A gas must be kept in an enclosed space when experiments are to be made upon it with regard to its alterations of volume under varying conditions of temperature. But we know that, so long as the temperature is maintained at a constant value the volume and pressure of the gas are in inverse proportion to each other. Hence we may investigate the effect of variation of temperature in two ways: we may measure the expansion under constant pressure; or, we may measure the change of pressure at constant volume.

By such measurements Charles (and, subsequently, Gay Lussac) was led to the conclusion that *the volume of a given quantity of any gas, under constant pressure, increases by a constant fraction of its amount for a given increment of temperature.* This statement is known as *Charles' Law*, and is represented symbolically in conjunction with Boyle's Law by the equation

$$pv = C\,(1+at),$$

where c and a are constants, and the meaning of the other quantities is obvious.

The pressure being kept constant, the volume increases by the fraction a of its amount at 0° per unit rise of temperature; a is therefore the co-efficient of dilatation under constant pressure. Again, if the volume be kept constant, the pressure increases by the fraction a of its amount at 0° per unit rise of temperature. Thus the fractional increase of volume under constant pressure, and the fractional increase of pressure at constant volume, have the same numerical value if the above equation be rigidly true.

The magnificent series of experiments carried out by Regnault have shown that, while the above law is very nearly true in the cases of the most permanent gases, marked discrepancies are exhibited when the more readily liquefiable gases are employed. His experiments on the undernoted gases gave as the value of the dilatation between 0° C. and 100° C.:

Hydrogen	0·3667	0·3661
Air	0·3665	0·3670
Nitrogen	0·3668	0·3670
Carbonic oxide ...	0·3667	0·3669
Carbonic acid	0·3688	0·3710
Cyanogen	0·3829	0·3877
Sulphurous acid ...	0·3845	0·3903

The figures in the second column represent the value of the coefficient as determined at constant volume; those in the third column are the values determined under constant (atmospheric) pressure. With a single exception (in the case of hydrogen) the latter number exceeds the former.

When the initial pressure of gases at 0° is increased, the dilatation between 0° and 100° increases. Some of Regnault's results for air are:

109·72 ...	149·31 ...	0·3648
374·67 ...	510·35 ...	0·3659
760·00 ...	1038·54 ...	0·3665
1692·53 ...	2306·23 ...	0·3680
3655·66 ...	4992·09 ...	0·3709

The numbers in the first two columns represent respectively the pressure per unit area at 0° and at 100° expressed in terms of the weight at 0° of a column of mercury of unit section and one millimetre in height.

For carbonic acid Regnault gives the similar results:

758·5 ...	1034·5 ...	0·36856
901·1 ...	1230·4 ...	0·36943
1742·9 ...	2387·7 ...	0·37523
3589·1 ...	4759·0 ...	0·38598

The variation in the case of this gas is therefore greater than in the case of air.

The results, for the same two gases, under constant pressure, are:

Air	760 ...	0·36706
	2525 ...	0·36944
Carbonic acid ...	760 ...	0·37099
	2520 ...	0·38455

Regnault's apparatus consisted essentially of a glass bulb D (Fig. 155), which contained the gas and communicated through the tube E with a reservoir AB which was filled with mercury.

A tube F, open to the atmosphere, also communicated with the reservoir. When the co-efficient was to be determined at constant volume, a plug C was screwed in until the mercury stood at the points E and F in the tubes—E being a fixed point. D was, under these conditions, surrounded successively by melting ice, and by the steam from boiling water, and the pressure in each case was found

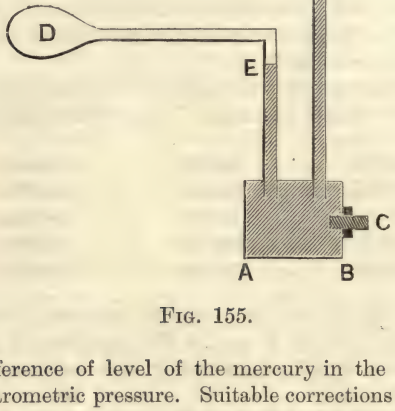


FIG. 155.

from the difference of level of the mercury in the two tubes and the known barometric pressure. Suitable corrections had, of course, to be applied for the expansion of the bulb by heating or by pressure.

When the dilatation was observed under constant pressure, the plug C was, at each temperature, screwed out until the mercury stood at the same level in the two tubes. The pressure was then equal to that of the atmosphere.

Other volumetric, or gravimetric, methods have been employed by Regnault and various experimenters.

266. *Absolute Zero of Temperature.*—We may write the equation which expresses Boyle's and Charles' Laws in the form

$$pv = \mathrm{C}\, a\left(\frac{1}{a} + t\right).$$

Now as the second term within the brackets represents temperature, the first term must also represent a temperature, for the dimensions (§ 27) of all the terms of a physical equation must be identical. And the numerical value of that temperature is about 273, since $a = 0{\cdot}003665$. Hence we must suppose that the zero of the Centigrade scale corresponds to a temperature of 273 degrees on this

new scale, which we may call a scale of *absolute* temperature, since its magnitude is independent of the particular gas employed.

To look at the matter from another point of view, we observe that, as t diminishes to zero and then becomes an increasing negative quantity, the product pv constantly diminishes, and finally becomes zero when $t = -273°$. If the volume is constant, this means that the pressure vanishes when t has this value. But the pressure of a gas is due to the motion of its particles, and hence, when $t = -273°$, the particles cease to move, and the gas is therefore totally deprived of heat.

The above equation may therefore be written in the form

$$pv = \mathrm{R}t,$$

where $\mathrm{R} = \mathrm{C}\alpha$ and t represents absolute temperature.

We shall see later (§ 303) that this estimate of the position of the absolute zero on the Centigrade scale is confirmed by thermodynamical considerations.

267. *Measurement of Temperature.*—The most usual method of measuring temperature is by means of the expansion of a liquid or a gas. Mercury is generally used in the former case, air in the latter.

A glass tube of narrow, and as nearly as possible uniform, bore is first chosen. If the bore be not quite uniform, its variations are determined by the process of calibration. This consists in running a small quantity of mercury along the tube from one end to the other, and measuring its length at the various parts. The quotient of the weight of the mercury by its specific gravity and the length of the column at any part gives the mean section of that portion of the tube. Next, a bulb is blown, on one end of the tube, of a size which is determined by the bore of the stem, the expansibility of the liquid to be used, and the required length of a scale division. The bulb is now slightly heated to expel some air, and the instrument is then inverted in a vessel of the liquid with which it is to be filled. As the bulb cools, some of the liquid enters it. This liquid is then boiled in the bulb, and its vapour expels the remaining air. A repetition of the process of inversion of the bulb and stem in the liquid will result in both being entirely filled. So much of the liquid is then run out that the remainder scarcely fills the stem when it is boiled once more. The vapour drives out the air which entered, and the tube is then hermetically sealed.

Two fixed points must now be determined on the stem. As we shall see afterwards (§§ 273, 275), this may be done by means of melting ice, and of the steam coming from boiling water in a nearly closed vessel.

But, before these points are determined, a considerable time should be allowed to elapse, for the bulb will not shrink quickly to its final volume. The process of shrinkage usually goes on for years, though, by careful annealing, the effect may be considerably lessened.

To determine the lower fixed point of the scale, the bulb and part of the stem are surrounded by melting ice. The final position of the extremity of the liquid column is marked on the stem.

The upper fixed point is determined by surrounding the bulb, and as much of the stem as possible, by the steam which is escaping from water boiling under a pressure of one atmosphere. The final position of the liquid is then marked on the stem.

The distance between the two marked points is then divided into a number of equal parts.

On the Centigrade scale, the lower fixed point is marked 0°, and the higher is marked 100°; on the Fahrenheit scale the lower is marked 32°, and the higher 212°. Thus, on the Centigrade scale, there are 100 divisions between the boiling-point and the freezing-point of water—hence its name; on the Fahrenheit scale there are 180 divisions between these points. The Fahrenheit zero was determined by a freezing-mixture of snow and salt, which gave the lowest temperature known at the time when Fahrenheit first constructed his thermometers. The relation between the two scales is obviously given by the equation

$$\frac{F-32}{180} = \frac{C}{100},$$

where F and C respectively represent the Fahrenheit and Centigrade scale readings.

The same interval on Réaumur's scale is divided into 80 equal parts, the zero being the same as that of the Centigrade scale.

As already remarked, the liquid generally used is mercury. For the measurement of temperatures below the freezing-point of mercury, alcohol is employed.

The air thermometer is of great use in the determination of temperatures above those at which mercury can be employed; and its readings agree pretty closely with those of the true absolute scale of temperature as determined by thermodynamical considerations. The following numbers, taken from Regnault's results, show the difference between the Centigrade scale of the air thermometer and that of the mercury thermometer:

Air ...	... 0°	20°	40°	60°	80°	100°	200°	300°
Mercury	... 0°	19°·98	39°·67	59°·62	79°·78	100°	202°·78	308°·34.

Self-registering thermometers are frequently employed for the purpose of indicating the maximum, or the minimum, temperature attained between two given periods. In the usual form of the maximum thermometer, the expanding column of mercury pushes an iron index along the tube. This index is left behind when the column contracts, for mercury does not wet iron. In the usual minimum thermometer, alcohol is used along with a glass index. The liquid, when expanding, flows past the index; when it contracts, it pulls the index with it, for its surface tends to take the smallest possible area (§ 120), and this is attained when it occupies the space between the index and the walls of the tube.

Continuously-registering thermometers are also used. The principle of one of the best forms of these instruments is identical with that of the Bourdon gauge. Let *abcd* (Fig. 156) be a longitudinal section of a hollow metal receiver in its unstrained state, and let us suppose that the receiver is filled with a liquid. As the temperature

FIG. 156.

rises, the pressure increases; and, since *cd* is greater than *ab*, the sum of the moments of the forces tending to straighten the receiver exceeds the sum of those which act contrariwise. A system of levers attached to such a receiver (fixed at one end) traces a continuous record on a properly-graduated paper placed on a drum which revolves slowly at a uniform rate.

Pyrometers are used for the rough determination of very high temperatures. In Daniell's pyrometer, a bar of platinum is slipped into a hole bored in a rod of graphite. A plug of graphite, or baked clay, rests on the top of the bar and fits tightly into the hole, or is otherwise kept tightly in position. When the platinum (which rests on the bottom of the hole) expands, the plug is pushed out, and remains in its position of maximum displacement when the temperature falls. From the increase of length of the platinum thus registered, we can calculate the temperature of any furnace in which it may be placed, *on the assumption* that the law of dilatation, determined throughout moderate ranges of temperature, holds up to the high temperatures of the furnace.

Other methods are also employed for the determination of high temperatures. See §§ 269, 343.

CHAPTER XXIII.

EFFECTS OF THE ABSORPTION OF HEAT: CHANGE OF TEMPERATURE AND CHANGE OF STATE.

268. *Unit of Heat. Specific Heat. Thermal Capacity.*—One of the most marked effects of absorption of heat is a rise of the temperature of the heated body. In some cases, no change of temperature takes place, and the effect which appears instead is a change of the physical state of the substance. But, before discussing these effects, we must consider the methods of measuring the amount of heat required to produce a given change; and this, in turn, necessitates the adoption of a definite unit of heat.

We may conveniently adopt as our unit *the quantity of heat which is required to raise the temperature of one pound of ice-cold water to* 1° C.

A given quantity of heat may therefore be measured by the number of pounds of ice-cold water which it can raise in temperature to 1° C. We might measure it also by the number of pounds of ice at 0° C. which it is just able to melt; for, as we shall see shortly, the number of units of heat which are required to just melt one pound of ice at 0° C. is quite definite and measurable. More generally, we might adopt as our unit the amount of heat necessary to produce any definite physical change, and then we might determine what fraction of the unknown quantity of heat could produce this change. The number of units in the unknown quantity would be the reciprocal of this fraction.

We define the specific heat of a substance, under given conditions, *as the quantity of heat which is required to raise the temperature of one pound of the substance by* 1° C. From this definition, and from our previous definition of the unit of heat, it follows that we must consider the specific heat of one pound of water at 0° C. to be unity.

More strictly, we should define it as the rate at which the quantity

of heat supplied to the substance, per pound of its mass, varies with the temperature. But, in all actual cases, there is no practical difference between the two definitions.

The *mean*, or *average*, specific heat of a substance, throughout a given range of temperature, is obtained by dividing the amount of heat which is required to raise one pound of the substance through the given range by the difference between the two extreme temperatures.

The *Thermal Capacity* of a substance is the quantity of heat which is required to raise the temperature of unit volume of the substance by one degree. It is therefore equal to the product of the specific heat and the density of the substance.

269. *Specific Heat of Solids and Liquids.*—Various methods are used for the determination of specific heat.

In one method, use is made of the fact that the rate of emission of heat from a body at a given temperature depends only upon the nature of its surface. Hence, if we fill a thin metal globe successively with two different liquids, and observe the rate of cooling of each liquid at the same temperature, we can compare the specific heats of the two liquids. For, if m, s, r, and m', s', r', represent respectively the mass, the specific, and the rate of cooling of the liquids, we have

$$msr = m's'r'.$$

If one of the liquids be water, so that $s = 1$, we get

$$s' = \frac{mr}{m'r'}.$$

In actual experiment, the liquids would be raised to a common high temperature and readings of their temperatures would be taken at equal intervals of time as they cooled. If a curve were then drawn the ordinates of which represented temperature, while the abscissæ represented time, the rate of cooling would be given by means of the tangent to the curve. Thus (Fig. 157), to find the rate of cooling at a temperature θ, draw the tangent ab to the curve at the point P, corresponding to θ, and let it intersect the time-axis in the point a and the temperature-axis at the point b; the rate of cooling is ob/oa.

Another method of determining specific heat is known as the *Method of Mixture.* Let m pounds of one substance at temperature t be mixed with m' pounds of another at temperature t', and let the specific heats of the two substances be respectively s and s'. If the temperature of the mixture be θ, the heat lost by the hotter body (say that of mass m') is $m's'(t' - \theta)$. Similarly, the heat gained

by the colder substance is $ms(\theta - t)$. Also, if μ be the mass of the vessel which contains the substances (and we may regard the stirring-rod used to mix the substances as forming part of it) while σ represents its specific heat, the heat given to the vessel is $\mu\sigma(\theta - t)$. Here we assume that the colder substance was originally contained in the vessel, the hotter substance being introduced from without.

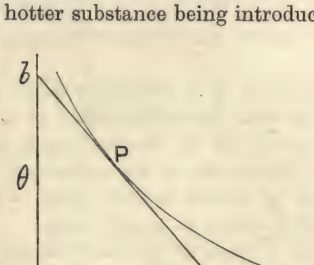


FIG. 157.

Instead of $\mu\sigma$ we may write w—a unit multiplier being understood. The meaning of this is that $w = \mu\sigma$ is the number of pounds of water (specific heat = unity) which require the same supply of heat to produce the rise of temperature $(\theta - t)$ as the vessel required. The quantity $\mu\sigma$ is therefore called the *water-equivalent* of the vessel.

We may therefore write

$$m's'(t' - \theta) = (m + w)(\theta - t),$$

provided that the cold substance is water. If the value of w be known, this equation enables us to find the value of s'. If w be not known, a second experiment in which the mass, m, of water is varied, will lead to another similar equation by means of which w may be eliminated or determined. Of course, the value of w may be found by one experiment alone, in which two quantities of water, at different temperatures, are mixed.

In an accurate experiment of this kind, precautions are taken that there shall be as little loss of heat by radiation as possible. This may be effected by making the vessel which contains the water (or other liquid) of a substance which is a bad radiator of heat; and, in addition, this vessel is placed inside a second similar vessel, contact between the two being prevented by means of bodies which do not readily conduct heat. If necessary, a third vessel may be used; and then a correction may be made for the slight amount of heat which is still lost by radiation.

A third method of determining specific heat is by the *Fusion of Ice*. As we shall see subsequently (§ 274), a definite amount of heat is required to just melt one pound of ice. Let H be this quantity. If M pounds of a substance melt m pounds of ice in the process of cooling from T° C. to 0° C., the average specific heat, S, throughout that range of temperature is given by the equation

$$m\mathrm{H} = \mathrm{MST}.$$

Bunsen and others have used forms of apparatus in which the quantity m is found by means of the decrease of volume of a mixture of ice and water when the heat which is given out by the cooling body melts some of the ice.

In general the specific heat of a substance increases with rise of temperature. But the specific heat of platinum varies very slightly with temperature, so that the range of temperature through which a mass of that metal cools is closely proportional to the quantity of heat which it emits. This fact is utilised in the measurement of high temperatures.

Regnault found that the quantity of heat required to raise the temperature of a pound of water from 0° C. to t° C. is represented by the equation

$$\mathrm{H} = t + 0{\cdot}00002t^2 + 0{\cdot}0000003t^3,$$

and that therefore the true specific heat, at any temperature, is given by

$$\frac{d\mathrm{H}}{dt} = 1 + 0{\cdot}00004t + 0{\cdot}0000009t^2.$$

His experiments were carried out at various temperatures between 0° C. and 230° C.

The specific heat of ice is almost exactly equal to 0·5, and Regnault found that it is diminished by decrease of temperature.

This table gives the specific heat of various elementary substances at ordinary temperatures :

	Solid.	Liquid.
Water	0·500 (at 0° C.)	... 1·000
Glass	0·180	
Iron	0·114	
Copper	0·096	
Zinc	0·094	
Silver	0·057	
Tin	0·056	... 0·064
Mercury	0·031	... 0·033
Lead	0·031	... 0·040

It is worthy of notice that the specific heat of water is considerably in excess of that of any of the other substances; and that, in general, the specific heat of any substance in the liquid state exceeds that of the same substance in the solid state.

270. *Law of Dulong and Petit.*—Dulong and Petit found that the product of the specific heat of any elementary solid into its atomic weight is practically constant. An alternative statement is that *the water-equivalent of an atom of each elementary solid is practically constant.* The numbers in the first column below represent specific heat; those in the second column represent atomic weight; and those in the third give the product of these two quantities.

Iron	0·114	54·5	6·2
Copper ...	0·096	63·5	6·1
Zinc	0·094	64·5	6·1
Silver	0·057	108·0	6·2
Tin	0·056	118·0	6·6
Mercury (liquid)	0·033	202·0	6·6
Lead	0·031	207·0	6·4

Similar results have been established for various series of chemical compounds. The value of the constant varies from one such series to another.

271. *Specific Heat of Gases and Vapours.*—The specific heat of a gas, at any given temperature, may be measured in two different ways. We may keep the pressure constant, or we may keep the volume constant. The numerical values of the specific heat, as obtained by the two methods, are different; and so we speak of the *Specific Heat at Constant Pressure* and the *Specific Heat at Constant Volume.*

The experimental difficulties which are encountered in the determination of the specific heat at constant volume are almost insurmountable; but, in the case of approximately perfect gases [which closely obey the law $pv = Rt$ (§ 266)], the principles of thermodynamics show (§ 301) that the difference of the two specific heats is equal to the quantity R. Hence it is only necessary to measure the specific heat of such substances at constant pressure. The method which is adopted consists in passing a slow stream of the gas, under constant pressure, through two spiral tubes, in the first of which its temperature is raised to a known amount, while in the second it is lowered to a known amount. The amount of heat which is given out in the process of cooling is measured by the rise of temperature of a known mass of water, and the mass of the gas which gives out the heat is determined from a measurement of its volume.

Delaroche and Bérard, who first used the above method, were led to believe that the specific heat of a gas varies with its pressure. Regnault, who worked with improved apparatus, found that it is independent of the pressure—not merely in the case of a gas such as air which sensibly obeys Boyle's law, but also in the cases of carbonic acid and hydrogen. He found also that while the specific heat of carbonic acid increases markedly as the temperature rises, the specific heat of air is independent of the temperature. It is, therefore, by Boyle's and Charles' laws, independent of the volume. And we may in all probability conclude that the specific heats of all gases which closely obey Boyle's law are absolutely constant. It follows that the thermal capacity of such a gas is proportional to its density. The results in the table are due to Regnault.

Specific Heat of Simple Gases.

Hydrogen	3·4090	Oxygen	0·2175
Nitrogen	0·2438	Chlorine	0·1210
Air	0·2374	Bromine	0·0555

Specific Heat of Compound Gases.

Ammonia	0·5084	Carbonic acid ...	0·2169
Carbonic oxide ...	0·2450	Hydrochloric acid	0·1852
Sulphuretted hydrogen	0·2432	Sulphurous acid ...	0·1544

The *ratio* of the two specific heats of a gas may be found from the speed of sound in that gas (see § 158). In air and some other gases the specific heat at constant pressure is almost exactly 1·4 times greater than that at constant volume. Jamin and Richard have obtained, by a direct experimental method (in which the temperature of the gas is raised by means of a known amount of heat developed by the passage of an electric current through a metallic wire), results which agree well with those obtained by the acoustic method.

The specific heat of water-vapour, under constant pressure, is about 0·48.

The specific heat of some saturated vapours—for example, those of water and carbon bisulphide—is *negative*. Such vapours, no liquid being present, become superheated under increase of pressure unless heat be withdrawn from them : and, under decrease of pressure, they will condense unless heat be supplied to them. Their specific heat diminishes in numerical magnitude as the temperature is raised. On the other hand, the specific heat of the saturated vapour of ether is positive, and increases with increase of temperature. In all cases

the actual increase of specific heat is positive. In benzine this increase results in a change of sign of the specific heat.

272. *Change of Molecular State. Latent Heat.*—We have already remarked that, in some cases, the application of heat to a body does not produce a rise of temperature, and that a change of molecular condition appears instead.

Thus the application of heat to ice below 0° C., raises its temperature and causes it to expand. When the ice reaches the temperature of 0° C., melting takes place continuously as more and more heat is applied. When the melting is complete the temperature again rises until, under ordinary atmospheric conditions, the water boils at 100° C. When all the liquid has boiled away the temperature again rises until the water begins to break up into its constituent elements.

If, at any stage of the above process, the application of heat were stopped, and heat were withdrawn instead, the various changes would be gone through in precisely the reverse order.

The change from the solid state to the liquid state is termed the process of melting or of *fusion*. The reverse process is called *solidification* or *regelation*.

The change from the liquid condition to the state of vapour is known as *vaporisation*, and the direct change from the solid state to the state of vapour is called *sublimation*.

In no case probably do these changes take place suddenly. Evaporation may go on at all temperatures; and many solids gradually soften before they melt. It is most probable that such softening occurs even in the case of ice and similar bodies which appear to melt suddenly.

The heat which is applied in order to produce fusion or vaporisation, without change of temperature, was called *Latent Heat* because it does not give rise to effects which can be measured by any ordinary thermometric apparatus.

273. *Fusion and Solidification.*—The laws which regulate the process of fusion and which have already been alluded to in the case of water, may be enunciated as follows:

1. *So long as the pressure is maintained constant there is a definite melting-point for every solid;*

2. *If the solid and the liquid be well mixed, and heat be applied slowly, the temperature of the mixture remains at the melting-point until the whole of the solid has melted.*

In the statement of the first law the condition of constant pressure is imposed. Only in the case of a substance, the volumes of equal masses of which, in the solid and liquid conditions

respectively, were equal, however the pressure might vary, would the restriction be unnecessary.

If a given liquid, such as water, expands in the act of solidification, the application of pressure will tend to prevent the solidification, *because* it tends to prevent expansion. Consequently, more heat must be withdrawn from the liquid in order that the change of state may be brought about. But this implies that the temperature is lowered.

Similarly, the melting-point (or rather, from our present point of view, the solidifying-point) of a substance, such as paraffin, which contracts in the act of solidification, is raised by the application of pressure. For the application of pressure makes the change occur more readily; and, consequently, less heat has to be abstracted in order that the action may proceed. In other words, the temperature at which the change occurs is raised.

The theoretical investigation of the problem will be given later (§ 300).

Professor James Thomson predicted from theory that the melting-point of ice would be lowered by pressure to the extent of 0°·0075 C. per atmosphere of pressure. This prediction was fully verified by Sir W. Thomson. The second column in the table below gives the melting-points of paraffin, in Centigrade degrees, which correspond to the pressures, in atmospheres, which are given in the first column. These results were obtained by Bunsen. The second and third pairs of columns give similar results, obtained by Hopkins, for stearine and for sulphur respectively.

1	... 46°·3	... 1	... 72°·5	... 1	... 107°
85	... 48°·9	... 519	... 73°·6	... 519	... 135°
100	... 49°·9	... 792	... 79°·2	... 792	... 140°

The motion of glaciers is due, in large part at least, to the fact that the melting-point of ice is lowered by pressure. When the pressure arising from the weight of the superincumbent strata of ice increases to a sufficient extent at any point in the bed of the glacier, liquefaction takes place, and the water flows round the obstacle to the presence of which the increase of pressure was due. But, the pressure being relieved at the given point (and, therefore, handed on to another part of the mass) because of the contraction which takes place in melting, the water again becomes solid: and so the glacier, by a continuous process of melting and re-solidification, gradually moves down the valley which it occupies.

For the same reason, snow which is not too cold may be readily kneaded into a compact mass of ice: and a wire, which is loaded at

its two extremities, and is hung over a bar of ice, will gradually cut its way through the bar without actually dividing it into two parts; for, though the ice below the wire is melted by the pressure, the water which is produced flows round the wire and solidifies above it. The path of the wire through the clear ice can be readily traced by means of the air-bubbles which the ice contains.

Sir W. Thomson has found that the earth as a whole is more rigid than an equal globe of glass. This could be explained if the melting-point of the average materials of the earth is raised by pressure. It is well known that this is so in the case of ordinary lavas.

Under special circumstances, the laws of fusion, as enunciated above, may be violated.

Thus Fahrenheit showed that water which completely fills a closed glass vessel may be cooled below 0° C. before it freezes. And Gay-Lussac showed that the temperature may be lowered to – 12° C. in an open glass vessel if the surface of the water be protected from the air by a layer of oil. The same phenomenon appears in the cases of other liquids, such as melted tin, phosphorus, and sulphur. In all such cases any vibration of the liquid must be avoided, or solidification will take place suddenly.

The melting-points of different substances vary greatly. On the one hand, hydrogen can only be solidified by the aid of powerful freezing mixtures; and, on the other hand, gas-coke can only be softened at the temperature of the electric arc.

Table of Melting-points.

Mercury... ...	−40°	Sulphur	115°
Ice	0°	Zinc	415°
Phosphorus ...	44°	Wrought iron ...	1500° (?)

The melting-point depends upon the purity of the substance. Thus the melting-points of different alloys of the same two substances vary greatly.

274. *Latent Heat of Fusion.*—The latent heat of fusion of any substance may be defined as *the quantity of heat which is required to just melt one pound of that substance at its ordinary temperature of fusion.*

This latent heat is given out again on re-solidification. Its amount, for each definite substance, under given conditions of pressure, is invariable.

The methods used for the determination of latent heat are essentially similar to those used for the determination of specific heat.

De la Provostaye and Desains, and Regnault, found the latent heat of fusion of ice to be equal to 79·25 units. Person, more recently, has used a different method, ultimately with the same result. He heated a quantity of ice, the temperature of which was originally below 0° C.; and, in consequence, he had to take account of the specific heat of ice, which, as we have seen, is about 0·5°. He at first obtained the value 80 for the amount of latent heat of fusion; but, subsequently, he traced the discrepancy between his result and that of previous observers to the fact that latent heat seems to be absorbed to a slight extent before the temperature 0° is reached. This, if true, furnishes evidence of the truth of the supposition, made in § 78, that the process of liquefaction is gradual.

The following values of the latent heat of fusion of some substances are taken from Person's results:

Latent Heat of Fusion.

Ice	79·25	Tin	14·25
Phosphate of soda ...	66·80	Lead	5·37
Zinc	28·13	Mercury	2·83

275. *Evaporation and Condensation.*—The laws of evaporation are similar to those of fusion.

1. *So long as the pressure is maintained constant, there is a definite boiling-point for every liquid.*

2. *If the liquid be well stirred, the temperature of both liquid and vapour remains at the boiling-point until all the liquid has evaporated.*

The effect of pressure is always in one direction with regard to the boiling-point, for all substances expand when they evaporate. The effect is therefore to raise the boiling-point, and its elevation is much more marked than is the alteration of the melting-point of a substance. But, before discussing this point farther, we must consider more fully the process which is termed *boiling*.

Evaporation occurs to a greater or less extent at all temperatures, and the rate of evaporation, *ceteris paribus*, increases rapidly as the temperature rises. If the liquid be contained in a closed vessel, the rate of evaporation gradually decreases and finally vanishes. (We assume, of course, that the area from which evaporation takes place remains constant. Under given conditions, the total rate of evaporation is proportional to the magnitude of the area.) It is not really true that evaporation has ceased. A state of kinetic equilibrium, in which the rate of evaporation is equal to the rate of condensation, has been attained. When this condition of equilibrium holds, the

vapour is said to be *saturated;* and it is found that the pressure of the saturated vapour depends only on the temperature.

The presence of gases, such as air, has no influence upon the final state of equilibrium: it merely increases the time necessary for the attainment of the condition.

But the vapour may be saturated at any temperature as well as at the usual boiling-point. And this leads to the definition of the boiling-point as *the temperature at which the pressure of the saturated vapour is equal to that to which the free surface of the liquid is subjected.*

The following remarks should make the matter clear. Let ABCD (Fig. 158) represent a cylinder in which a smooth, massless (and therefore weightless) piston AD, which we also suppose to be gas-tight, works freely. First, let there be a gas in the closed region P and another gas in the region Q outside the

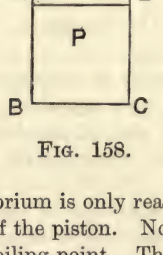


FIG. 158.

piston. Evidently equilibrium is only reached when the pressure is the same on both sides of the piston. Now suppose that P is filled with a liquid below its boiling-point. The filling of the region P is necessarily complete so long as vapour is not formed. And no vapour can be formed until the pressure of that vapour is equal to the pressure of the gas in the region Q, *i.e.*, until its pressure is equal to the pressure to which the free surface of the liquid is exposed. But when the liquid is at the temperature at which this occurs, vapour will be formed, and the continued application of heat will force the piston up. If we now suddenly produce a vacuum in the region Q, vapour will be rapidly formed in P; and that vapour will proceed not merely from the surface of the liquid but also in bubbles from its interior. This process of free evaporation is called *boiling* or *ebullition.*

[The phenomena exhibited by Geysers are due to a like cause. A sudden reduction of pressure in the interior of the column of water

which fills the funnel causes the water at that part to change its state explosively, and so the superincumbent water is ejected violently.]

The following table, given by Regnault, exhibits the relation between the boiling-point and the pressure:

Pressure of the Saturated Vapour of Water.

Temp. C.	Pressure in Atmospheres.	Temp. C.	Pressure in Atmospheres.
0°	0·006	120°	1·962
10°	0·012	130°	2·671
20°	0·023	140°	3·576
30°	0·042	150°	4·712
40°	0·072	160°	6·120
50°	0·121	170°	7·844
60°	0·196	180°	9·929
70°	0·306	190°	12·425
80°	0·466	200°	15·380
90°	0·691	210°	18·848
100°	1·000	220°	22·882
110°	1·415	230°	27·535

By sufficiently reducing the pressure, water may be made to boil violently—not merely to evaporate at its surface—at temperatures far below its ordinary boiling-point. The well-known experiment of causing hot water to boil in a closed flask, by pouring cold water upon the flask, is a case in point. The sudden reduction of temperature causes partial condensation of the vapour already formed in the flask, and so gives rise to a sudden diminution of pressure.

When the boiling-point is known, the atmospheric pressure may be obtained from a table such as that above. The *Hypsometric Thermometer*, used for the determination of height above sea-level, is based upon this principle. The atmospheric pressure diminishes as the elevation above sea-level increases, and the result is that the boiling-point is lowered by about 1° C. at an elevation of 960 feet above sea-level.

The laws of evaporation are subject to exceptions, just as are the laws of melting. Thus water may, by cautious heating in a smooth clean glass vessel, be raised considerably above its ordinary boiling-point, if it has been carefully freed from dissolved gases. A very slight vibration may then cause it to boil explosively.

The boiling-points of various liquids differ greatly under ordinary atmospheric conditions, as the following table shows:

Table of Boiling-points of Liquids.

Zinc	1040° C.	Bisulphide of carbon ...	48° C.
Mercury	350°	Sulphurous acid	−10°
Water	100°	Nitric oxide	−87°

276. *Latent Heat of Vaporisation.* — Regnault found for the 'total heat of steam,' *i.e.*, the quantity of heat which is given out by one pound of water in condensing to water at 0° C., the expression,

$$H = 606{\cdot}5 + 0{\cdot}305t,$$

where t is the temperature in Centigrade degrees. This gives for the latent heat the expression

$$L = H - \int_0^t \sigma dt,$$

where σ is the specific of water. The value of σ is (§ 269) $1 + 0{\cdot}00004t + 0{\cdot}0000009t^2$. If we substitute this value in the integral, we get

$$L = 606{\cdot}5 - 0{\cdot}695t - 0{\cdot}00002t^2 - 0{\cdot}0000003t^3.$$

This formula is true throughout the range of temperature from 0° C. to 230° C. If we could assume that it held true up to 706° C., it would indicate that the latent heat vanishes at that temperature very nearly. (See § 278.)

The latent heat of steam is very large in comparison with that of most other liquids, as this table shows:

Latent Heat of Vaporisation.

Water	536	Ether	90·4
Naphtha	264	Bisulphide of carbon	86·7
Alcohol	202	Bromine	45·6

We have found previously that the latent heat of liquefaction of ice is also relatively large. These facts—of the large latent heats of liquefaction of ice and of vaporisation of water—are of great importance in the economy of nature. If they were not so, destructive floods might frequently occur from the rapid liquefaction of ice, or sudden condensation of moisture, consequent on a slight variation of temperature.

The latent heat of vaporisation is used for the production or maintenance of low temperatures. Water may be kept cool in very

hot weather if it is enclosed in a vessel of porous earthenware; for part of it percolates through the vessel and evaporates from its outer surface, the latent heat being largely drawn from the vessel and its liquid contents. Also solid carbonic acid is produced if a jet of the liquid (formed under considerable pressure) is allowed to escape from the vessel which contains it; for the outer parts of the jet evaporate, and the necessary latent heat is largely taken from the interior parts of the jet, which consequently are solidified. Faraday froze mercury in the interior of a white-hot platinum crucible, by placing it in a capsule which rested on a mixture of solid carbonic acid and ether contained in the crucible. Similarly, in very hot countries, ice may be formed at night on shallow pools because of rapid evaporation.

277. *Formation of Dew.*—When a superheated vapour is cooled sufficiently, saturation takes place, and any further cooling causes condensation. The moisture which is deposited in this way from the atmosphere is termed *dew*. Any cold body lowers the temperature of the air in immediate contact with it; and, when the temperature is sufficiently lowered, a thin film of moisture is deposited upon the cold body. The latent heat which is given out on condensation gradually raises the temperature of the cold body until it becomes equal to that which corresponds to the vapour-pressure, at which stage the action ceases. This temperature, being also that at which the deposition of dew just commences, is called the *Dew-point*. *Hoar-frost* is formed when the dew-point is below 0° C.

Wells first gave the correct explanation of the formation of dew. He showed that dew is freely deposited on nights when the sky is clear, because on such nights the earth loses heat rapidly by radiation and so cools rapidly to the dew-point; whereas, on cloudy nights, the clouds absorb, and radiate back to the earth, a large part of the radiated heat, so that the ground does not cool rapidly. Another condition necessary to the ready formation of dew is that the air shall be still, otherwise no portion of the air may remain in contact with the ground for a length of time sufficient to allow of its being cooled to the dew-point. The dew will, of course, deposit itself most freely on those bodies which part with their heat most rapidly and have also small specific heat.

Aitken has recently shown that the presence of particles of dust is necessary before condensation of moisture can occur in the atmosphere, and that supersaturation of a vapour can be produced by getting rid of all dust-particles by filtration through cotton-wool. These particles act as nuclei upon which the deposition takes place. This phenomenon is very closely connected with the phenomenon of

the dependence of the equilibrium-pressure of vapour upon the curvature of the liquid film with which it is in contact (§ 127).

Aitken has also utilised the fact that moisture is deposited upon the dust-particles in the construction of an instrument which enables us to determine the number of particles which are contained in a given volume of any definite specimen of air. This instrument is certain to prove of considerable meteorological importance.

Daniell's *Hygrometer* was constructed for the purpose of accurately registering the dew-point. It consists of two hollow glass bulbs connected by a glass tube. One of these bulbs is made of black glass, and the other is made of clear glass. A small thermometer, the stem of which projects into the tube of (clear) glass which connects the two bulbs, is placed in the black bulb along with a quantity of sulphuric ether. The remaining portions of the interior of the instrument are filled only with the vapour of ether. A piece of cambric is tied round the other bulb, and a little ether is poured upon it. The evaporation of this ether cools the bulb, and makes some of the vapour inside it condense. This destroys the equilibrium of the liquid ether and its vapour, in the interior of the instrument; and some of the ether in the black bulb evaporates, in order that equilibrium may be restored. The absorption of latent heat cools this bulb, and, finally, dew is deposited on its exterior. The presence of a very slight film of dew is readily observed on the black surface, and the temperature of the bulb is noted; but the reading of the thermometer is necessarily a little too low. The evaporation is then stopped, and the temperature at which the dew just disappears is observed. This reading is a little too high, and so the mean of the two results is taken.

Regnault introduced improvements which rendered it possible to observe the appearance and disappearance of the dew at practically one temperature.

The dew-point is also found by means of *Wet and Dry Bulb Thermometers*. The one thermometer has its bulb surrounded by cambric, which is kept moist with water drawn up by capillary action through some threads which dip into a vessel containing it. The other (ordinary) thermometer registers the exact temperature of the air. The reading of the wet-bulb thermometer is lower than that of the dry-bulb thermometer so long as evaporation is going on; but, if the atmosphere is saturated with water-vapour, no evaporation takes place, and both thermometers register the same temperature. The formula

$$p = p_0 - \frac{\delta}{48} \cdot \frac{b}{30},$$

in which p represents the pressure of water-vapour in the atmosphere, p_0 represents the pressure which is given in Regnault's table as corresponding to the temperature of the wet-bulb, δ is the difference between the wet-bulb and the dry-bulb readings, and b is the height of the barometric column in inches, was found by Apjohn to accord well with observed results.

278. *Continuity of the Liquid and Gaseous States. Critical Temperature.*—Cagniard de la Tour first showed that a substance may exist in a non-liquid state at a density very nearly equal to its density in the liquid condition. A complete investigation of the subject was made by Andrews, who showed that

There is a Critical Temperature for every vaporous or gaseous substance, such that no amount of pressure can liquefy the substance, unless its temperature be below the critical value.

The critical temperature of carbonic acid is 30°·9 C. That of water is about 412° C.

The latent heat vanishes at the critical temperature. We have already seen that the latent heat of water should vanish at about 706° C., if Regnault's formula connecting latent heat with temperature held throughout that range. The result just given shows that the formula deviates largely from the truth at temperatures higher than the limit (230° C.) up to which Regnault worked.

The accompanying diagram (Fig. 159) represents the results of Andrews' experiments on carbonic acid. Pressure is measured (in atmospheres) along the axis of ordinates, and volume is measured along that of abscissæ. At the temperature 13°·1 C., the volume of the gas gradually diminishes, as the pressure is raised, until liquefaction commences. After this, the volume lessens, without any rise of pressure, until all the substance is liquefied; and then immense pressure is required to lessen it even slightly.

Similar effects take place at the higher temperature 21°·5. The line (called an *isothermal*) which represents the simultaneous values of pressure and volume at this higher temperature, lies, in the diagram, entirely to the right hand of, and above, the isothermal of 13°·1; for, the pressure being constant, the volume increases with the temperature, and, the volume being constant, the pressure increases with the temperature. But, at this higher temperature, the change of volume in passing from the gaseous to the liquid state is smaller than that which occurs at the lower temperature; and liquefaction commences at a smaller volume, and ends at a larger volume, than when the temperature is less. The isothermals cease to have a portion parallel to the axis of volume, *i.e.*, liquefaction ceases, at 30°·9.

The dotted curve separates the region in which the liquid and the vapour can exist together in equilibrium from the regions in which the substance is entirely liquid or entirely vapour. The isothermal of 30°·9 separates the region in which liquefaction can occur from that in which it is impossible.

[Compare, with this diagram, Fig. 11, which is drawn, so as approximately to suit the case of carbonic acid, from theoretical considerations based, by Professor Tait, upon the kinetic theory of gases.]

We may with great advantage, as Tait suggests, describe the substance as a *true gas*, or a *true vapour*, according as the temperature is higher, or lower, than the critical temperature.

The compressibility of the substance is dv/vdp, where v is the volume and dv, dp, represent respectively simultaneous small increments of the volume and the pressure. Now, the diagram shows that, at the commencement of liquefaction, the inclination of the isothermal to the axis of volume becomes greater and greater as the temperature rises, *i.e.*, the ratio dv/dp decreases as the temperature rises. Hence, since we suppose unit volume to be taken in all cases, the compressibility of the vapour when it is upon the point of condensing decreases as the critical temperature is approached. Similarly, the value of dv/dp in the liquid state, when the substance has just been entirely liquefied, increases as the temperature rises; and thus we see that the compressibility in the two states tends towards equality, simultaneously with the volumes, as the temperature rises to its critical value.

For a considerable distance above the critical point the isothermals exhibit two points of inflexion; but these finally cease to be visible, and the isothermals closely resemble those of a perfect gas.

[The illustration affords a good example of the use of contours. The isothermals may, as was stated in Chap. III., be regarded as the projections of the plane sections of a surface which represents the various simultaneous values of the pressure, volume, and temperature of the gas.]

279. *Solution. Freezing Mixtures.*—The process of solution is extremely analogous to the processes of liquefaction. A gas which is dissolved in a liquid may be regarded to a certain extent as if it were liquefied, and latent heat is given out in the process of solution. Similarly, when a solid is dissolved in a liquid, latent heat is required, just as if the solid were directly liquefied; but in some cases this absorption of latent heat is masked by the heat which is developed because of molecular action between the liquid and the solid.

The amount of heat which is disengaged in the solution of a gas

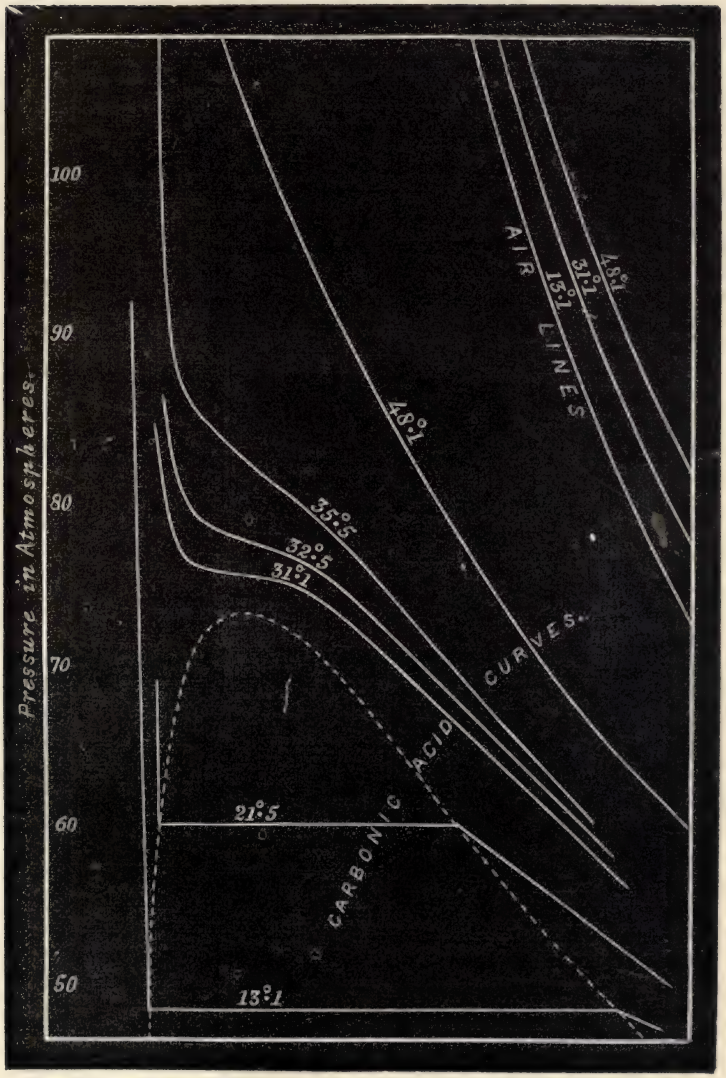

Pressure in Atmospheres.
100
90
80
70
60
50
AIR LINES.
48°1
31°1
15°1
48°1
35°5
32°5
31°1
CARBONIC ACID CURVES.
21°5
13°1

FIG. 159.

is frequently very marked. This is so specially in the cases of the more soluble gases, such as ammonia when water is the solvent.

The amount of gas, under definite pressure, which a given liquid will dissolve, becomes less as the temperature is raised; though, by careful treatment, a state of supersaturation may be induced—which is analogous to the prevention of boiling at temperatures considerably over the ordinary boiling-point of a liquid.

Supersaturation of a liquid solution of a solid may also take place—notably in the case of a substance, such as acetate of soda, which dissolves in little more than its own water of crystallisation. If a crystal of the acetate be dropped into the supersaturated solution to act as a nucleus, crystallisation will take place rapidly with the development of latent heat. A crystal of any other substance of the same crystalline form will produce the same effect.

Heat is frequently developed or absorbed when two liquids are mixed (mutually dissolved). If chemical action takes place to any extent between the two, heat will be developed unless other causes prevent. If the total bulk of the two liquids increases on mixture—as in the case of bisulphide of carbon and alcohol—heat tends to be absorbed; and again, the water-equivalent of the mixture may be greater than the sum of the water-equivalents of its constituents—which also necessitates absorption of heat. If the opposite effects to these take place, heat will be evolved. Energy may also be changed into heat in the process of inter-diffusion of the liquids. Disengagement or absorption of heat take place respectively according as the effects of the one or the other sets of actions preponderate. And, as the various actions depend upon the temperature, we find that the total effect is sometimes reversed when the original temperature of the two liquids is sufficiently varied.

Two solids even may dissolve in each other, so to speak, with the absorption of latent heat. (Salt and snow furnish a well-known example.) This can obviously only occur when the freezing-point of the resultant liquid is lower than the original (common) temperature of the solids. Part of the latent heat is obtained by cooling the solids, part by cooling the liquid, and part, it may be, by cooling surrounding bodies. If the whole be intimately mixed the action necessarily ceases when the freezing-point of the resultant liquid is reached. These remarks contain the explanation of the action of solid freezing-mixtures, which has been elaborately investigated by Professor Frederick Guthrie.

280. *Dissociation and Chemical Combination.*—When the temperature is raised sufficiently high a compound dissociates, or breaks up, into its constituents. The change is not sudden but gradual.

It commences at a certain lower limit of temperature, and ends completely at a certain higher limit; and at all intermediate temperatures a state of kinetic equilibrium is arrived at in which recombination precisely balances dissociation. The magnitude of the limits will, in general, depend upon the pressure. It is usual to speak of the temperature at which one-half of the substance is dissociated as the temperature of dissociation.

Conversely, when the two (or more) constituents are mixed, combination does not occur until a certain temperature is attained; but, if the combination results in the development of heat, the process, once started, will continue until the percentage of the mixture which remains uncombined corresponds to the temperature which the whole mass attains because of the heat which is set free. On the other hand, if work is done during the process, or if heat is lost by conduction or otherwise, the process will continue until combination is complete when the temperature falls to the lower limit.

All chemical combination takes place in accordance with the two laws of thermodynamics, and therefore further treatment of this subject is deferred until we have considered these laws. (See § 298.)

281. Many other effects of heat might be noted here, but it is preferable to leave their discussion to those special sections in which we have to treat of the properties which are affected by the application of heat.

CHAPTER XXIV.

CONDUCTION AND CONVECTION OF HEAT.

282. *Conduction.*—We have already discussed the transference of heat by the process of radiation, that is, the transference of heat without the mediation of ordinary matter. In the process of radiation the transferred energy may pass *through* a material substance without being communicated to it; indeed, radiation ceases in so far as such communication is made. We must now consider its transference when ordinary matter is the medium through which it is transferred.

The most marked difference between the two cases lies in the rate of propagation, which is extremely rapid when radiation occurs, while it is extremely slow in comparison when ordinary matter is the medium of transference.

Two methods exist according to which heat (which consists in kinetic energy of molecular motion) may pass from one place to another by means of matter. The energy may pass from one portion of matter to another, which occupies a different position, by actual (or virtual) impact between the molecules of the two portions; that is to say, it may be handed on from one portion to another: or again, it may pass, not from one portion of matter to another, but from one locality to another by motion of the hot body. The former of these processes is known as *Conduction;* the latter as *Convection.* Both take place in liquids and in gases; the former alone can take place in solids.

283. *Conductivity.*—Different substances, under like conditions, conduct heat at different rates. Thus a bar of iron, one end of which is red-hot, may be too hot to grasp at the cooler end; while a bar of wood, of the same length, which is burning at one end, may be easily handled at the other. The property in virtue of which such differences arise is termed *Conductivity.*

Most experiments which are intended to illustrate the differences between the conductivities, or conducting-powers, of various sub-

stances for heat exhibit only the differences between the rates at which the temperatures of the substances, at a given distance from the source of heat, attains a definite value under given conditions. The well-known experiment of Ingenhouz is of this description. In it, a series of similar and equal rods, of different substances, project from the side of a metallic trough into which hot water is suddenly poured. Each rod is coated with a thin film of beeswax which melts at a definite temperature. The rate at which this definite temperature travels along each rod is plainly shown by the motion of the line of demarcation between the melted and the unmelted portions of the wax. But obviously this rate will only coincide with the rate at which heat is conducted along when the thermal capacities of the various substances are practically identical, for, other things being equal, the rate at which the temperature rises is inversely proportional to the thermal capacity.

Fourier was the first to give an accurate definition of conductivity. The whole subject of heat-conduction was so fully and accurately developed by him that his work—'Théorie analytique de la Chaleur'—published in 1822, still remains the text-book on the subject.

Let us suppose that the substance, the thermal conductivity of which we are considering, is in the form of a uniformly thick plane slab of practically infinite extent. Let θ be its thickness; and let one side be kept at uniform temperature t, while the other is kept at uniform temperature t' until a steady flow of heat takes place from side to side. The quantity of heat, h, which passes in τ units of time through an area a of the surface of the slab is found experimentally to be directly proportional to τ, a, and $t'-t$, while it is inversely proportional to θ. Hence we may write

$$h = k\, a \frac{t'-t}{\theta} \tau.$$

The quantity $(t'-t)/\theta$ is called the *temperature gradient*, and k is the conductivity.

If the area, the temperature gradient, and the time, be all unity, the equation becomes

$$h = k,$$

and so we obtain the following definition of the conductivity:

The thermal conductivity of a substance, at any temperature, is the number of units of heat which pass, per unit of time, through unit of surface of an infinite slab of the substance, of unit thickness, the sides of which are kept respectively at temperatures half a degree higher, and half a degree lower, than the given temperature.

In making this definition we assume that the unit of length is not excessively small—that it is (say) a centimetre, an inch, or a foot—and that an ordinary temperature degree—say, the Centigrade—is used, so that the temperature gradient is not large. The necessity for these restrictions is apparent, if we consider that the conductivity may (it actually does) vary somewhat with the temperature; for, in consequence of such variation, the temperature gradient could not be sensibly uniform from side to side of the slab, if the difference of the temperatures at the two sides were large. As a special case, let us suppose that the conductivity of a layer of the slab, of half its total thickness, is one-half of that of the remaining portion. Since the same flow of heat takes place through both portions, the difference of temperature between the sides of the former portion must be double of that between the sides of the latter.

Of course, even if the conductivity varies from point to point, whether from variation of temperature or from any other cause, the quantity k, determined from the above formula, will always represent the average conductivity of the slab considered as a whole.

But, quite apart from the question of such variation, we cannot assert that the quantity of heat which will pass through a slab, one unit in thickness, under unit difference of temperature, will be precisely equal to the quantity which will pass through a slab, n times thinner, under a difference of temperature n times less, when n is a very large number, and all the other conditions are unaltered.

284. *Measurement of Conductivity.*—In one form of experiment for the absolute determination of conductivity, a steady state of temperature is maintained throughout the substance. This method was used by Lambert, and subsequently, under greatly improved conditions, by Forbes.

In Forbes' method a long bar of the substance of uniform cross-sectional area is used. One extremity of the bar is inserted in a bath of melted lead, or solder; and the other extremity is exposed to the air, or, if necessary, is cooled by a current of water. Small holes, into which a little mercury is poured, are drilled in the bar at regular intervals; and these holes are lined with iron (if the bar itself be not made of iron) in order to prevent amalgamation. Thermometers, inserted in the holes (which are found not to appreciably affect the flow of heat along the bar) register the temperature of the bar in their immediate vicinity.

If distance measured along the bar from the source of heat be laid off along ox (Fig. 160), and if ordinates be drawn at points such as p, corresponding to the positions of the thermometers, and of lengths which are proportional to the readings of the thermometers at these

points, a curve drawn free-hand through the extremities of the ordinates will enable us to obtain the temperature gradient at any

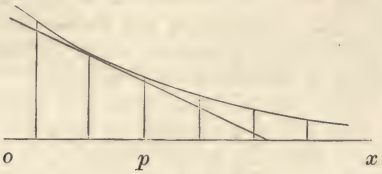


FIG. 160.

part of the bar. For the tangent of the angle of inclination of the line which touches the curve at the extremity of any ordinate is equal to the space rate at which the temperature varies, per unit of length, at the corresponding section of the bar; *i.e.*, it is equal to the gradient of temperature at that section. But the sectional area of the bar is known, and hence, if we can determine the quantity of heat which passes in a given time through the given section, we can determine the conductivity by calculation from the equation above.

Now the heat which passes any section is entirely lost from the remaining portion of the bar by radiation, or otherwise; and any heat which is given to the water employed in cooling the far end of the bar, if this is required, can be readily estimated by means of the rise in temperature of the water, while the heat which is lost by radiation and convection is found by a special experiment.

During the above experiment, a thermometer, inserted in a hole in a small bar which is cut originally from the long bar, indicates the temperature of the air in the neighbourhood of the large bar. In the second experiment, which is made for the purpose of determining the rate of loss of heat, the small bar is heated uniformly to a temperature higher than the highest recorded in the former one. The bar is now allowed to cool, and the thermometer which is inserted in it enables us to determine the rate of loss of heat per unit of time, per unit of length of the bar (§ 269). The mass of unit length is known, the specific heat is also determined, and the product of these quantities into the rate of fall of temperature gives the rate of loss of heat. This being known for all the various temperatures observed at the different parts of the bar in the first experiment, the total rate of loss of heat from the portion of the large bar, beyond any given section, is easily calculated. And so the conductivity, at particular temperatures, can be found.

In the second experiment the temperature of the air is obtained

by means of the long bar, so that the results in both cases can be compared, as is necessary, at the same excess of temperature over that of the surrounding air. But in addition to this, since the rate of cooling depends upon temperature and pressure of the air, it is necessary to perform both experiments under as nearly as possible the same conditions of temperature and pressure.

The unit of heat which is employed in this method is obviously the amount of heat which is required to raise the temperature of unit volume of the substance by one degree, for the amount of the heat is measured in terms of changes of temperature in the bar. Consequently the quantity which is so determined is not the therma conductivity as above defined. Maxwell calls it the *Thermometric Conductivity;* Thomson calls it the *Thermal Diffusivity.* The thermal conductivity of any substance is obviously the product of the thermometric conductivity into the thermal capacity of that substance.

Tait, who repeated and extended Forbes' experiments, gives the following values of

Thermometric Conductivity.

Temperature C.	0°.	100°.	200°.	300°.
Iron	0·0149	0·0128	0·0114	0·0105
Copper, electrically good ...	0·076	0·079	0·082	0·085
Copper, electrically bad ...	0·054	0·057	0·060	0·063
German silver	0·0088	0·009	0·0092	0·0094

This table indicates that, with the exception of that of iron, the thermometric conductivity of all these substances increases as the temperature rises.

Tait also gives, for the iron and the two specimens of copper (the units being the foot, the minute, and the degree C.), the following values of

Thermal Conductivity.

Iron	$0{\cdot}788\ (1-0{\cdot}00002t)$
Copper, electrically good ...	$4{\cdot}03\ \ (1+0{\cdot}0013t)$
Copper, electrically bad ...	$2{\cdot}84\ \ (1+0{\cdot}0014t)$

It appears, therefore, that in general the thermal conductivity increases as the temperature rises. Also, the order of the metals with regard to conduction of heat is the same as their order with regard to conduction of electricity. Forbes had observed this fact, and had expected that, as in the case of electric conduction, the thermal conductivity would decrease as the temperature becomes higher. This, as we see, does not, in general at least, hold true.

Dr. A. C. Mitchell has recently, under Professor Tait's direction, repeated these experiments with the same bars, but under improved conditions. For one thing, all the bars were nickel-plated, so as to avoid alterations of the surface from oxidation at high temperature. His results are, on the whole, confirmatory of the previous results —with this chief exception, that he found the temperature co-efficient for iron to be positive, as it is in all the other substances.

A different method was employed by Angström. In his method one end of the bar is alternately heated and cooled during equal periods of time, the temperature of the source of heat being kept constant. The alternations of heating and cooling are maintained until all the thermometers indicate practically periodical changes of temperature. Fourier's mathematical investigations show that, if the variations of temperature do not sensibly affect the conductivity and the specific heat, the conductivity can be calculated from the rate at which the range of temperature diminishes per unit of length of the bar, together with the observed speed at which the 'waves of temperature' run along the bar, provided that the rate of surface-loss of heat is proportional to the excess of the temperature of the bar over that of its surroundings.

285. *Conduction through the Earth's Crust.* — Angström's method has a direct application to the problem of the conduction of the diurnal and annual waves of solar heat downwards through the crust of the earth. In this investigation we may assume that the heated surface is practically an infinite plane, and that the flow of heat takes place in lines perpendicular to this plane.

Let ab (Fig. 161) represent the surface, and let cd, ef, represent planes parallel to the surface, at distances x and $x+\delta x$, respectively, from it. If c be the thermal capacity of the substance through which

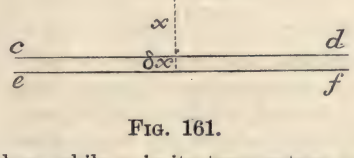


Fig. 161.

the flow takes place, while v is its temperature, and t represents time, the quantity of heat which enters a portion of the substance, of small thickness δx and area a, in a small interval of time δt is

$$ca\frac{dv}{dt}\delta t\delta x, \quad \ldots\ldots (1)$$

for $ca\delta x$ represents the quantity of heat which must be abstracted in order to lower the temperature of the volume $a\delta x$ of the substance by one degree, and $dv/dt \,.\, \delta t$ is the change of temperature in the time δt.

If k be the conductivity of the substance, the quantity of heat which, in the time δt, crosses in the positive direction the area a of the side of the slab nearest the surface is

$$-a\delta t k\frac{dv}{dx}.$$

Similarly, the quantity which passes downwards through the area a of the surface which is distant from the former by the amount δx is (since δx is small)

$$-a\delta t\left(k\frac{dv}{dx}+\frac{d}{dx}\left(k\frac{dv}{dx}\right)\delta x\right).$$

Consequently, the amount of heat which, on the whole, enters the volume $a\delta x$ in the time δt is

$$a\delta x\delta t\frac{d}{dx}\left(k\frac{dv}{dx}\right), \quad . \quad . \quad (2)$$

and hence we get

$$c\frac{dv}{dt}=\frac{d}{dx}\left(k\frac{dv}{dx}\right), \quad . \quad . \quad (3)$$

since each of the quantities (1) and (2) represents the same amount of heat.

Now (§§ 27, 67) let us consider this equation simply as an equation of dimensions. The difference of temperature dv appears linearly on both sides of (3), and therefore the range of temperature does not appear in the dimensional equation. We get

$$\frac{c}{t}=\frac{k}{l^2}, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where t represents time, and l represents length; for dx, the dimensions of which are those of a length, appears twice as a factor in the denominator on the right-hand side of (3); while k and dt appear once only on the right-hand side and the left-hand side, respectively, of that equation; and we must remember that the sign of equality indicates now equality of *dimensions* alone.

From (4) we get

$$l=\sqrt{\frac{kt}{c}},$$

which means that, if the times are altered in any fixed proportion p, the lengths must be altered in proportion to the square root of p in order that the flow of heat may take place under *similar* conditions in the altered circumstances. In other words, *the distances at which similar effects are felt* (*for example, the distances below the surface at which the periodic variations of surface-temperature cease to be felt*) *are proportional to the square root of the period.*

Now the period of the annual heating and cooling is 365 times as great as the period of diurnal variations. Hence the effect of the summer's heat is felt about nineteen times as far below the surface as the effect of the diurnal heat is felt.

Again,

$$\frac{l}{t} = \sqrt{\frac{k}{ct}}.$$

Here we may suppose that l represents the length of a wave, while t is the periodic time, so that the fraction on the right hand is proportional to the rate at which the wave of heat travels downwards. We see, therefore, that this rate is directly proportional to the square root of the conductivity, and is inversely proportional to the square root of the thermal capacity and the periodic time conjointly.

It follows that, when the period is constant, *the date at which the maximum temperature reaches any given depth is later than the date at which it left the surface in direct proportion to the depth.*

The law which regulates the diminution of the range of temperature with increase of depth cannot be obtained from equation (3) in the way in which we have obtained the two laws just enunciated; for the temperature does not appear in equation (4). But we may write (3) in the form

$$c\frac{dv}{dx}\frac{dx}{dt} = \frac{d}{dx}\left(k\frac{dv}{dx}\right);$$

and we may suppose that dx/dt is the speed with which the heat-wave travels downwards, in which case the equation becomes

$$\sqrt{\frac{ck}{\mathrm{T}}} \cdot \frac{dv}{dx} = \frac{d}{dx}\left(k\frac{dv}{dx}\right),$$

where T is the periodic time, and dv/dx now represents the rate at which (say) the maximum temperature changes as the wave passes down.

This equation asserts that the rate of diminution of the rate of change of temperature with depth is proportional to the rate of change itself. In other words, the rate of change diminishes in

geometrical progression as the depth increases in arithmetical progression. And its rate of diminution is $\sqrt{c}/\sqrt{k}\mathrm{T}$. But, since the rate of diminution of the rate of alteration of the range is proportional to the rate of alteration itself, it follows that the rate of alteration bears the same ratio to the range. Hence *the range diminishes in geometrical progression as the depth increases in arithmetical progression*, the rate of diminution being directly as the square root of the thermal capacity, and inversely as the square roots of the conductivity and the periodic time conjointly.

As the result of direct observations (begun by Forbes in Edinburgh in 1837) of the temperature at different distances below the surface of the earth, it is found that the annual heat-wave travels inwards at the rate of little more than sixty feet per annum, and that the range of temperature has diminished to a very small fraction of its original amount when half of that distance has been traversed. The diurnal heat is therefore inappreciable at a depth of at most about two feet. Of course, all these results depend upon the nature of the soil.

The thermometers nearest the surface are affected by changes in the weather, but these disturbances rapidly die out.

When a steady state of temperature is reached, dv/dt vanishes, and (3) becomes

$$\frac{d}{dx}\left(k\frac{dv}{dx}\right)=0.$$

This gives

$$k\frac{dv}{dx}=a \text{ (constant)},$$

which shows that the temperature-gradient varies inversely as the conductivity. This applies directly to the case of the earth regarded as a cooling body, and shows us that *in strata throughout which the conductivity does not vary, the temperature increases uniformly per unit of depth.* Of course, if the earth is regarded as a *cooling* body, the steady state of temperature is impossible; but its rate of cooling is so slow that the time-variations of temperature may be neglected.

Fourier's equations, when applied to past time as regards the earth or the sun, indicate a state of uniform high temperature throughout the mass—a state which could not have arisen by any process of conduction. This suggests the production of the heat by the gravitation of separate masses (§ 93).

286. *Conduction in Crystalline Bodies.* — Crystalline bodies possess, in general, unequal conducting power in different directions.

The conducting power is symmetrical with regard to three rectangular axes—called the principal axes of thermal conductivity.

If a single point-source of heat were placed in the interior of a crystal, the loci of constant temperature would be concentric ellipsoids surrounding that point; and Stokes has shown that the conductivities parallel to the axes of these ellipsoids are proportional to the squares of the axes. Sections of such an ellipsoid can be obtained by means of thin plane plates of the crystal cut in different directions from the substance. If a copper wire be passed through a small hole drilled through such a plate in the direction of the axis of the ellipsoid conjugate to the plane, and if this wire be heated by an electric current, the heat conducted away by the plate may be made to melt a thin coating of beeswax on the surface of the plate. The boundary between the melted and the unmelted wax is a section of the above ellipsoid.

287. *Conduction in Liquids and Gases.*—The thermal conductivities of liquids (neglecting liquid metals) are small in comparison with those of solids; and those of gases are smaller still.

In experimental investigations on the subject, great care must be taken to avoid convection currents (see below), which would completely invalidate the results.

In the case of gases, the conductivity can be calculated from the kinetic theory.

288. *Convection.*—Under gravity, all liquids and gases tend to arrange themselves in horizontal layers the densities of which decrease as their distances from the earth's surface increase. This condition may be entirely disturbed because of variations of temperature; for the consequent changes of density destroy the equilibrium, and currents are set up in the fluid so as to restore it. These are called 'convection-currents.'

We have already discussed a typical case in dealing with the maximum density point of water.

Very marked examples occur on a large scale in nature. The trade-winds are due to the ascent of hot air currents in equatorial regions, while colder air blows in from the polar regions to take its place. [The north-easterly or south-westerly direction of these winds is due to the rotation of the earth.] A considerable part, at least, of ocean circulation, is also of the nature of convection. Again, whenever water evaporates, heat is absorbed to be evolved wherever condensation occurs. This is indeed one of the most fruitful sources of violent storms; for, if sufficient heat is developed, the consequent increase of temperature causes a rapid up-rush of air, and so creates

a partial vacuum which gives rise to a violent inflow of the surrounding air. [The air which comes from the south has (in the northern hemisphere) a greater eastward motion than that which comes from the north, and so a counter-clockwise vortex motion is produced. Thus the rotation of a cyclone is explained.]

Practical applications of the principles of convection are seen in the usual methods of boiling water, of ventilation, etc.

A hot body, which is cooling in a gas or a liquid, loses heat by convection as well as by radiation. The law of convective cooling in a gas has been elaborately studied by Dulong and Petit. Their results are expressed by the formula

$$r = ap^b\theta^{1.233},$$

where r is the rate of cooling, a is a constant for a given gas and a given body, p is the pressure of the gas, b is a constant for any one gas, and θ is the excess of the temperature of the cooling body over that of the gas. The rate is independent of the nature of the surface of the body but varies with its form and dimensions.

CHAPTER XXV.

THERMODYNAMICS: HEAT AND WORK.

289. *Mechanical Equivalent of Heat. First Law of Thermodynamics.*—Heat, since it is a form of energy, may be transformed into mechanical work and into all other forms of energy; but the former transformation is the only one with which we are at present concerned.

Colding and Joule were the first, after Rumford, to make determinations of the amount of work which can be produced from a given amount of heat, *i.e.*, of the *mechanical* (or, more properly, the *dynamical*) *equivalent of heat.*

The most direct method of conducting such an investigation consists in spending a known amount of work in the production of heat by friction. This method was used by Joule, who caused a falling weight to drive a vane rotating in the interior of a calorimeter which contained a known amount of water. The amount of heat developed was determined by means of the observed increase of the temperature of the water, due precaution being taken to correct for the loss of heat by radiation, and for the heat developed by friction between the parts of the apparatus. Various other methods were used by Joule, Hirn, Regnault, and others. For example, Joule proved experimentally that the heat developed by the sudden compression of air is practically equivalent to the work spent in compression (§ 294); and from this result, together with an accurate determination of the specific heats of air, he obtained the value of the mechanical equivalent. He also determined its value by means of the heat developed in a conductor by the passage of a current of electricity through it under given conditions (§ 342). One of Hirn's series of experiments was made upon a heat-engine in actual use. In another series, he experimented upon the heat developed by percussion. The latter of these gave a good result; the former did not.

Joule finally gave the number 772 (in foot-pounds at the latitude

of Manchester) as the amount of work necessary to raise the temperature of one pound of water by one degree Fahrenheit. The equivalent of the heat-unit (Centigrade) which we have hitherto used, is therefore 1,390 foot-pounds.

These experiments prove the law of conservation of energy in so far as heat and work are concerned. The statement of the law of conservation for these two forms of energy goes by the name of the FIRST LAW OF THERMODYNAMICS, which asserts that *when equal quantities of mechanical effect appear from purely thermal sources, or disappear in the production of thermal effects alone, equal quantities of heat disappear or are produced.*

290. *Carnot's Complete Cycle of Operations*—Although, as is indicated in the previous section, we can determine, by experiment, the direct relation between given amounts of heat and work, we are not entitled to draw any conclusion regarding the relation between the heat which disappears and the work which appears in any given physical process, unless a certain condition be observed. The necessity for this condition (which has already been referred to in § 254) was pointed out by Sadi Carnot.

The condition is that the working substance must pass through a complete cycle of operations, *i.e.*, a cycle at the end of which the substance has returned to its original physical state. Heat may have been expended in the given cycle, and work may have been produced; but, unless the final state is the same as the initial state, we cannot say that the work and the heat are mutually equivalent. For example, carbonic acid gas is heated by compression, but the heat developed is not the equivalent of the work spent in compression, for work is done by the molecular forces during the process.

In order to be able to reason correctly upon the connection between heat and work, Carnot assumed the existence of a heat-engine which can never be realised in practice. But this does not render his results any the less valuable, for, in order that we may be able so to modify Carnot's results that they may apply to the special case, we only require to know in what way, and to what extent, any given engine differs in its action from Carnot's.

He assumed that his engine was furnished with a cylinder the sides and piston of which were absolutely impermeable to heat, while the bottom was a perfect conductor of heat. He assumed also the existence of two bodies, one hot and the other cold, the temperatures of which were kept constant. The former of these was to act as the source of heat; the latter was to act as the condenser. The working substance is supposed to be placed in the cylinder below the piston, and may be any substance whatsoever,

with any properties whatsoever. But, for the sake of definiteness, we may assume that it acts as the steam in an ordinary engine would act.

Let us suppose that the cylinder, with its contents at the temperature of the cold body, is placed on a non-conductor of heat. The contents will retain their temperature (t_0, say) so long as the cylinder remains on the non-conductor, for the working substance is now surrounded on all sides by non-conductors. Let the volume and pressure of the substance be denoted by v_0 and p_0 respectively. [Remark here that the physical condition of a known mass of the substance is completely determinate when any two of its temperature, volume, and pressure are given.]

As the first operation of Carnot's cycle, the cylinder still remaining on the non-conductor, press down the piston until the temperature of the substance rises to that of the hot body (t_1, say) and the volume and pressure become v_1 and p_1 respectively.

As the second operation of the cycle, place the cylinder, with the condition of its contents unaltered, on the hot body, and let the substance slowly expand until its volume becomes (say) v_2, and the pressure becomes p_2 ($< p_1$). The expansion must occur so slowly that heat can flow into the substance so as to constantly prevent its temperature from being finitely different from t_1.

As the third operation place the cylinder and its contents—again without variation of condition—upon the non-conductor, and let the working substance expand until its temperature falls to that of the cold body. Let the pressure and the volume now be p_3 ($< p_2$) and v_3 respectively.

Lastly, place the cylinder upon the cold body and slowly press down the piston—the temperature of the contents remaining at t_0 until the volume again becomes v_0, and the pressure, therefore, again takes the value p_0.

This series of operations obviously satisfies Carnot's condition that the final condition of the working-substance shall be identical with its initial condition, *i.e.*, it forms a complete cycle.

Now in the second operation heat was taken from the hot body, and, in the fourth operation, heat was given to the cold body. Let these quantities be h_1 and h_0 respectively.

Also, in the first and fourth operations, work was done upon the contents in diminishing their volume; and, in the second and third operations, the contents, in their expansion, did work against the external pressure. Let these quantities be w_0 and w_1 respectively. In the first and fourth operations the volume decreased from v_3 to v_1, and the temperature was either equal to t_0 or rising from t_0 to

t_1: in the second and third operations the volume increased from v_1 to v_3, and the temperature was either equal to t_1 or falling from t_1 to t_0. In the former pair the temperature was therefore on the whole lower than it was in the latter pair. Consequently the pressure was higher when work was being performed by the substance than when it was being expended upon it, and therefore $w_1 - w_0$ is positive. And, the cycle being complete, we can write

$$J(h_1 - h_0) = w_1 - w_0,$$

where J is the multiplier (the mechanical equivalent, sometimes called Joule's equivalent—hence the letter J) required to change the heat units into dynamical units. This equation is the analytical expression of the First Law of Thermodynamics.

291. *Carnot's Reversible Cycle.*—Besides the idea of a complete cycle of operations Carnot introduced the equally important and fruitful idea of a *Reversible Cycle.* This is a cycle which can be performed in the exact reverse order: and Carnot's cycle can be so performed.

First. Begin with the cylinder on the non-conductor, the temperature, volume, and pressure of its contents being t_1, v_1, and p_1 respectively. Let the substance expand until these quantities become t_0, v_0, and p_0.

Second. Place the cylinder on the cold body and let the expansion proceed until the pressure and volume become p_3 and v_3 respectively, the temperature being still t_0.

Third. Place the cylinder on the non-conductor, and push down the piston until the temperature rises to t_1, the pressure and volume becoming p_2 and v_2 respectively.

Fourth. Place the cylinder on the hot body and compress the contents until the initial conditions are again attained.

Now in the first and second of these reverse operations work was done by the substance, while the temperature had either the low value t_0 or was falling from t_1 to t_0: and, in the third and fourth operations, work was done upon the substance, while the temperature remained at the high value t_1, or rose from t_0 to t_1. On the whole, therefore, the temperature—and consequently the pressure—had a higher value when work was done upon the substance than when work was performed by it. But heat was absorbed from the cold body in the second operation, and was given to the hot body in the fourth. In the reverse cycle, therefore, heat has been pumped up from the condenser to the source of high-temperature heat, but

work has been expended in the process. The complete action is also represented by the equation of last section.

Carnot reasoned upon the supposition that heat was material and so believed that the quantity of heat which was absorbed from the hot body was given to the cold body in the direct process, and that the quantity which was absorbed from the cold body was given to the hot body in the reverse process. He supposed that the heat did work in the direct process merely in being let down from a source at high temperature to a sink at low temperature, just as water does work in falling from a high to a low level.

The interpretation of his result, on the principle of conservation of energy, is simply that the excess of the heat absorbed over that emitted is directly transformed into its equivalent in work; and that, in the reverse operation, the heat-equivalent of the work expended, together with the heat absorbed from the cold body, is equal to the heat given to the hot body.

292. *Reversibility the Test of Perfection. Second Law of Thermodynamics.*—We shall now discuss one of the many important results which can be deduced from Carnot's principles.

Let an engine be reversible in the sense that all its physical and mechanical actions can be performed in the exact reverse order. Such an engine is perfect in the sense that it is as perfect as any engine, working under the same conditions, can be, *i.e.*, the reversible engine, taking in a quantity h_1 of heat at temperature t_1, and working with its condenser at the temperature t_0, will perform as much work as will any other engine, working through the same range of temperature, and also taking in the quantity of heat h_1 at the temperature t_1.

Carnot proved this statement by showing that, if it were not true, the perpetual motion would result. Let M denote the reversible engine, and let N denote the (supposed) more perfect engine. Make N work directly between the temperatures t_1 and t_0, taking in a quantity of heat h_1, giving out a quantity h_0, and performing an amount of work $W > w$ the quantity of work which the reversible engine would produce under the same conditions. Make M work backwards between the same sources at the same high and low temperatures. A quantity of work w is all that has to be expended in order to make it take in the amount h_0 of heat at the temperature t_0 and restore the amount h_1 to the source at the high temperature. Hence an excess of work $W - w$ is gained in the double process, while, on the whole, no heat has been transferred either way. But this means the perpetual production of work from nothing, which is impossible. Hence the reversible engine is perfect.

To adapt this reasoning of Carnot to the modern ideas of energy, we have only to argue thus: Work is performed in the double cycle, while no heat is on the whole taken from the source, therefore N must give less heat to the cold body than M takes from it, and so the double engine can only work by giving to the hot body heat which it has taken from a colder body, the temperature of which it constantly lowers; but this is in opposition to all known facts, and the denial of its possibility may be safely taken as axiomatic.

This adaptation is due to Sir W. Thomson, who re-introduced Carnot's work to the scientific world when it had been long disregarded, and who applied his principles to the deduction of thermodynamical results of the highest importance.

The statement that an engine which is reversible, in the sense that all its physical and mechanical actions are capable of exact reversal, converts, under given conditions, the greatest possible fraction of the heat which is supplied to it into useful work is known as the SECOND LAW OF THERMODYNAMICS. We shall find in § 298 that its analytical expression is $\Sigma(h/t) = o$.

293. *Absolute Temperature.*—The *efficiency* of a heat-engine is the ratio of the quantity of heat which it utilises in the form of mechanical work to the total quantity of heat which is supplied to it. In the notation used above it is

$$\frac{h_1 - h_0}{h_1}.$$

The value of this fraction is a maximum, under any given conditions, when a reversible engine is used. And *all* reversible engines, whatever may be the nature and properties of the working-substance, are equally perfect (*i.e.*, they possess the same efficiency) when they work through the same range of temperature. This enables us, as Thomson pointed out, to obtain an *absolute* measure of temperature—absolute in the sense that it does not depend upon the properties of any particular substance.

The definition which Thomson finally adopted was framed so as to make the absolute scale coincide as nearly as possible with the scale of the air-thermometer. It is this:

'*The temperatures of two bodies are proportional to the quantities of heat respectively taken in and given out in localities at one temperature and at the other, respectively, by a material system subjected to a complete cycle of perfectly reversible thermodynamic operations, and not allowed to part with or take in heat at any other temperature; or, the absolute values of two temperatures are to one another in the proportion of the heat taken in to*

the heat rejected in a perfect thermodynamic engine, working with a source and refrigerator at the higher and lower of the temperatures respectively.'

Expressed in symbols, this gives

$$\frac{t_1}{t_0} = \frac{h_1}{h_0},$$

where t_1 and t_0 are the absolute temperatures of the source and the condenser, respectively.

294. *The Indicator Diagram.* — The indicator diagram was introduced by Watt for purely practical purposes. It is, nevertheless, as we shall shortly see, susceptible of numerous important applications in pure science.

The diagram exhibits the relation between the pressure and the volume of the working substance used in any heat engine; and hence it shows the work done in a complete stroke of the engine.

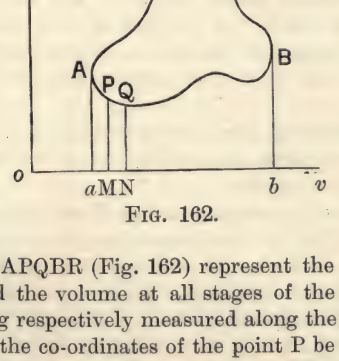


FIG. 162.

Let the curve APQBR (Fig. 162) represent the relation between the pressure and the volume at all stages of the stroke—pressure and volume being respectively measured along the rectangular axes op and ov. Let the co-ordinates of the point P be PM$=p$, OM$=v$; and let those of Q be QN$=p'$, ON$=v'$. When P and Q are indefinitely near each other, the product $(p+p')(v'-v)$ represents twice the area PMNQ. But, neglecting small quantities of the second order, half this product may be written $p(v'-v)$. This latter product therefore represents the elementary area PMNQ.

Now the work done by a force f, acting through a distance s, is (§ 62) fs; and the whole force which acts upon the piston in the cylinder of the engine is pa, where a is the area of the piston, and p is the pressure per unit area, so that pas represents the work done

when the piston moves through the small distance s under the action of the pressure pa. But as is the small change of volume of the contents of the cyclinder which is produced by that pressure; and therefore the work done is represented by the area PMNQ.

Let A be the point at which the volume has its smallest value, v_0, and let B be the point of maximum volume, v_1. As the volume expands from v_0 to v_1, the state of the substance being represented by a point which moves along the path APQB, the work performed is represented by the area AabBQPA. Similarly, when the volume is diminished from v_1 to v_0, the point moving from B to A along the path BRA, the work expended in producing compression is represented by the area BRAabB. The difference of these two areas, viz., the curvilinear area APQBRA, therefore represents the work which is expended, on the whole, during a complete stroke of the engine. Consequently, when the path is described in the positive direction, work is expended on the whole. This case corresponds to the reverse working of a reversible engine.

When the closed path is described in the negative direction, work is performed on the whole by the engine to an extent which is represented by the total area enclosed by the path. [The actual path might consist of a number of closed loops. In this case, each loop is to be considered separately, and the sum of the areas—each with the proper sign attached, according as it is described positively or negatively—is to be taken in order to estimate the total amount of work performed.]

295. *Applications of the Indicator Diagram.*—We shall now consider the application of the diagram to the discussion of the working of Carnot's engine.

For this purpose we must give the closed curve a special form. In the second and fourth direct operations of Carnot's cycle (§ 290), the substance was maintained at constant temperature; in the first and third, no heat, *as such*, was allowed to pass out of it or to enter it. If, therefore, the points D, A, B, C (Fig. 163) represent respectively the values (p_0, v_0; p_1, v_1; p_2, v_2; p_3, v_3) of the pressure and volume of the substance at the commencement of the first, second, third, and fourth operations respectively, all points on the line DA (which represents the varying state of the substance as it passes from the state (p_0, v_0, t_0) to the state (p_1, v_1, t_1) are characterised by the condition that no heat, as such, enters or leaves the working substance; all points on the line AB represent states in which heat is absorbed during the passage from the condition A to the condition B in order that the temperature may retain the value t_1; all points on the line BC represent states in which, again, no heat enters or

leaves the substance; and all points on the line CD represent states in which the temperature of the substance is, by disengagement of heat, maintained at the value t_0.

Lines such as AB or CD are therefore called *isothermals*, while lines such as DA or BC are called *adiabatics*. The latter name, which is due to Rankine, simply implies that no heat is absorbed or emitted.

But although heat, *as such*, neither enters nor leaves the substance in the adiabatic condition, mechanical work is performed or

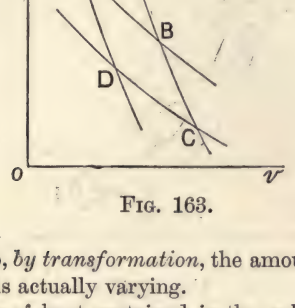


FIG. 163.

expended; and so, *by transformation*, the amount of heat contained in the substance is actually varying.

If the amounts of heat contained in the substance, in the states A and D, were identical, and if the amounts in the states B and C were identical, the amounts of heat which leave the substance in the processes represented by AB and DC would necessarily be equal. But we know that these quantities are (by our definition) in the ratio

$$\frac{h_1}{h_0} = \frac{t_1}{t_0}, \quad \ldots\ldots (1)$$

t_1 and t_0 being absolute temperatures.

Now ABCD represents the work which is performed in the direct cycle. It therefore, by the principle of conservation of energy, represents the quantity of heat $h_1 - h_0$. And, if we arrange matters so that $h_1 - h_0 = 1$, equation (1) shows not only that $t_1 - t_0$ must be equal to unity, but also that $h_1 = t_1$, and $h_0 = t_0$.

Let us therefore intersect the diagram (Fig. 164) by a series of isothermals $A_1A_2A_3\ldots$, $B_1B_2B_3\ldots$, etc., and by a series of adiabatics $A_1B_1C_1\ldots$, $A_2B_2C_2\ldots$, etc.; and let the temperature corresponding to $B_1B_2B_3\ldots$ be one degree below that corresponding to $A_1A_2A_3\ldots$,

while the adiabatics are so arranged that the magnitude of each area, such as $A_1A_2B_2B_1$, is unity. If we could continue this construction down to the absolute zero of temperature, each area included between any isothermal, the isothermal of absolute zero, and any two consecutive adiabatics, would be numerically equal to the temperature indicated by the higher isothermal.

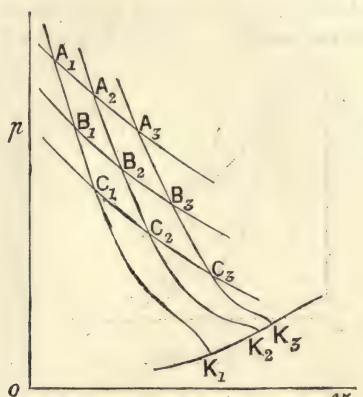


FIG. 164.

The experimental data necessary to the correct completion of the diagram are wanting, but its correct completion is not necessary to our present purpose. The following method of completing it is due to Maxwell.

Let us suppose that $C_1C_2C_3 \ldots$ is the lowest isothermal whose form is correctly known. Draw any line $K_1K_2K_3 \ldots$ to represent the isothermal of absolute zero, and complete the adiabatics in such a way that each of the areas $C_1K_1K_2C_2$, $C_2K_2K_3C_3$, etc., is numerically equal to the temperature to which the line $C_1C_2C_3 \ldots$ corresponds. In order to determine the position of the absolute zero, it is only necessary to find the ratio of any two areas such as $A_1K_1K_2A_2$ and $C_1K_1K_2C_2$. The method by which Thomson and Joule solved this problem is described in § 303.

296. *Applications of the Indicator Diagram. Entropy.*—The one set of lines with which we have intersected the diagram are characterised by constancy of temperature. The other set of lines are characterised by the condition that no heat shall be transferred from the working substance to its surroundings, or from the sur-

roundings to it. But we have seen that heat does pass from, or into, the substance by transformation. We have still to inquire, therefore, What quantity remains constant during adiabatic expansion or contraction?

Equation (1) of last section enables us to answer the inquiry. It gives

$$\frac{h_1}{t_1} = \frac{h_0}{t_0},$$

whatever values t_1 and t_0 may have; that is to say, when the substance passes isothermally from any one definite adiabatic state to any other definite adiabatic state, the quantity of heat which is absorbed, or disengaged, bears a definite ratio to the temperature at which the absorption, or disengagement, takes place.

This constant quantity h/t may therefore be regarded as the amount by which some quantity, ϕ, which is characteristic of the substance in the adiabatic state, changes in passing from the one adiabatic condition to the other at constant temperature t. This suggests the extension of the meaning of the quantity h to signify the total heat contained in the substance, so that we may define ϕ by the equation

$$\phi = \frac{\mathrm{H}}{t},$$

where H represents the total heat.

The quantity ϕ was first called, by Rankine, the *Thermodynamic Function*. Clausius called it the *Entropy*, and this name has been generally adopted.

If one substance parts with an amount of heat, h, at temperature t, to another substance at temperature t_0, the entropy of the former substance decreases by the amount h/t_1, and that of the latter increases by the amount h/t_0. The total gain of entropy by the system is therefore

$$h\left(\frac{1}{t_0} - \frac{1}{t_1}\right).$$

The quantity of heat which is absorbed in the second operation of Carnot's direct cycle is $t_1(\phi_1 - \phi_0)$, and the quantity which is given out in the fourth operation is $t_0(\phi_1 - \phi_0)$. The difference of those quantities is $(t_1 - t_0)\ (\phi_1 - \phi_0)$, which therefore represents the work done in the complete cycle.

297. *Applications of the Indicator Diagram. Total, Available, and Dissipated Energy.* — We have no means of determining experimentally the total amount of energy in a given system. But,

in any practical case, we only require to determine the *change* of energy which takes place in the given operations.

The indicator diagram enables us to represent the *total* amount of energy in a way which, though obviously incorrect, leads to a correct representation of the *change* of the total energy of the system which is produced by a given change in its physical condition.

Let the points A and B (Fig. 165) represent respectively the initial and the final conditions of the system, and let the path AB represent the series of changes by which the final condition was arrived at. [The diagram is constructed so as to exhibit the case in which there

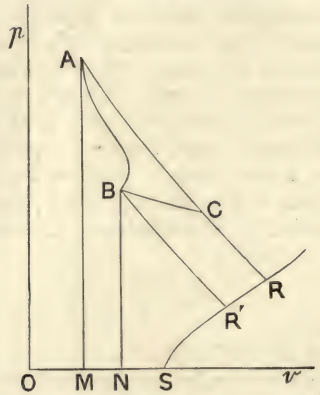


FIG. 165.

is both disengagement of heat and performance of mechanical work.] Let RR′ represent the (arbitrary) isothermal of absolute zero, and let it be continued so as to cut the axis of volume in the point S.

As the point which traces out the diagram moves from A to B, external work, which is represented by the area ABNMA, is performed, and at the same time, § 295, an amount of heat is disengaged which is represented by the area ABR′RA. The whole area AMNBR′RA, therefore, represents the total loss of energy which the working substance has sustained. We may, therefore, as Maxwell suggested, regard the areas AMSRA and BNSR′B, respectively, as representing the total amounts of energy which are contained in the working substance in the conditions indicated by A and B respectively.

It is specially to be noticed that the amount of energy which is lost in proceeding from A to B is totally independent of the path AB, and depends only on the initial and final conditions of the substance.

Let us assume now, as a special case, that BC represents the lowest available temperature. We may then cause the substance to pass from the state A to the state B along the path ACB. The work performed will then be represented by the area ACBNMA, which area therefore represents the total amount of energy which is available, under the given conditions, for the performance of mechanical work. Similarly, BCRR′B represents the heat which is necessarily given to the condenser, *i.e.*, the amount of energy which is necessarily dissipated so far as the performance of work by the given system is concerned. The energy which is unnecessarily dissipated is represented by the area ACB.

298. *Thermodynamic Motivity.*—We have seen that, when a quantity of heat, h_1, is given out by a body at temperature t_1, the entropy diminishes by the amount h_1/t_1. We may therefore represent the total loss of entropy of a system which consists of a number of sources which are emitting heat at various temperatures by the symbol

$$\Sigma\left(\frac{h_1}{t_1}\right).$$

Similarly, if the heat emitted by these bodies is given to other bodies included in the same system, we may denote the gain of entropy from this source by

$$\Sigma\left(\frac{h_0}{t_0}\right).$$

The total loss of entropy is therefore

$$\Sigma\left(\frac{h_1}{t} - \frac{h_0}{t_0}\right).$$

Deleting the suffixes, we may denote this simply as

$$\Sigma\left(\frac{h}{t}\right),$$

and our definition of absolute temperature shows that *this vanishes when a perfect engine is used.* When any other engine is used, heat is always lost by conduction or otherwise, so that the heat which is given to the condenser (or is otherwise wasted) is greater than h_0. Hence, in all actual cases of the transformation of heat into work we must write

$$\Sigma\left(\frac{h_1}{t_1} - \frac{kh_0}{t_0}\right),$$

instead of

$$\Sigma\left(\frac{h_1}{t_1} - \frac{h_0}{t_0}\right),$$

where k is greater than unity, as the proper expression for the loss of entropy. This is equal to

$$\Sigma\frac{h_1}{t_1}(1-k),$$

which is necessarily *negative*, since k is greater than unity; and so we prove Clausius' theorem that *the entropy of the universe tends to a maximum.*

The amount of heat utilised by a perfect heat-engine is (§ 293)

$$h_1 - h_0$$

$$\text{or } h_1 - \frac{t_0}{t_1}h_1.$$

If a number of such engines work between the various parts of a complex system, the heat which is not given to the condenser is

$$\Sigma(h) - t_0\Sigma\left(\frac{h}{t}\right),$$

where $\Sigma(h)$ now takes account both of the heat which is taken from each body by some of the engines, and of the heat which is given to it by others.

In a perfect engine, the heat which is not given to the condenser is represented by $\Sigma(h)$ alone, and is entirely converted into work; which again shows us that, for such an engine,

$$\Sigma\left(\frac{h}{t}\right)$$

must vanish. In all other cases, $\Sigma(h)$ represents the part of the heat which is utilised in the performance of mechanical work, and the second term (which we must remember is necessarily positive) represents the portion which is unnecessarily wasted. The quantity

$$t_0\Sigma\left(\frac{h}{t}\right)$$

therefore represents the heat which is dissipated in the process.

Thomson has called the total energy which could be made available for mechanical work by a perfect engine under given conditions, the *Thermodynamic Motivity* of the system. If we have a medium external to the given system, which may be used as a condenser, the motivity is the whole amount of work which can be obtained by the perfect engine in reducing the temperature of the system to that of the external medium. If the engine works so as to equalise the temperatures of the various parts of the system, the motivity is the whole amount of work which can be so obtained.

Let t_0 be the final temperature, and let h be a quantity of heat taken from a body at temperature t. The motivity, so far as this quantity is concerned, is

$$h\frac{t_1-t_0}{t_1};$$

and, to obtain the total motivity, we must sum all such quantities.

The total energy, e, in any system is equal to the sum of the motivity, m, and the dissipated energy of that system. If ϕ is the entropy, θ the temperature, and J the mechanical equivalent of heat, the dissipated energy is $J\theta\phi$. Hence, if the system passes from a state indicated by the suffix 1 to a state indicated by the suffix 2, we get

$$m_1 - m_2 = e_1 - e_2 - J(\theta_1\phi_1 - \theta_2\phi_2).$$

No change can take place of *itself* in the system unless thereby the motivity be decreased—that is, unless $m_1 - m_2$ be positive. But $m_1 - m_2$ may be positive, although $e_1 - e_2$ is negative, provided that $\theta_2\phi_2$ is sufficiently greater than $\theta_1\phi_1$. Hence we see that a given chemical action may take place of itself with absorption of heat, provided that a sufficient amount of energy be dissipated in the process. (See § 280.)

CHAPTER XXVI.

THERMODYNAMICAL RELATIONS.

299. In the course of the discussion, in last chapter, of the connection between heat and work, we were led to consider five quantities in terms of which the physical condition of a substance may be represented. These quantities were the energy, e; the entropy, ϕ; the pressure, p; the volume, v; and the temperature, t.

But we were also led to see that the physical condition of the substance was completely determined when two of these quantities, p and v, were given: for, on the indicator diagram, we could lay down lines of constant temperature, of constant entropy, and of constant energy. The *total* values of the two latter quantities were not, it is true, indicated; but that was due to a defect in our knowledge, and not to any defect necessarily inherent in the diagram.

It at once follows that the variation of any one of the five quantities can be represented in terms of the simultaneous variation of any two of the rest. Thus we may write

$$de = ad\phi + bdv, \ldots\ldots (1).$$
$$de = fdt + gdv, \ldots\ldots (2).$$
$$de = mdt + ndp, \ldots\ldots (3).$$

and so on.

300. We shall first consider equation (1). If the volume be constant we obtain $de = ad\phi$. But we know that, under constant volume, the energy increases by the amount (in dynamical units) of heat which has been supplied; and we also know, by the results of last chapter, that this amount is $td\phi$. Hence $a = t$. Similarly, if no heat be supplied, so that ϕ remains constant, (1) becomes $de = bdv$. But under these conditions the energy *diminishes* by the amount of external work which is performed, that is, by the amount pdv. Hence $b = -p$, and (1) becomes

$$de = td\phi - pdv \ldots\ldots (4).$$

This gives

$$\left(\frac{de}{d\phi}\right)_v = t, \left(\frac{de}{dv}\right)_\phi = -p.$$

The suffixes denote respectively that the volume and the entropy are constant; so that the former equation asserts that, at constant volume, the increment of the energy is equal to the heat supplied; while the latter asserts that, under adiabatic expansion, the decrement of the energy is equal to the amount of work which is performed. These equations therefore express the conditions upon which we deduced (4) from (1).

From them we get

$$\frac{d^2e}{dvd\phi}=\left(\frac{dt}{dv}\right)_\phi,\quad \frac{d^2e}{d\phi dv}=-\left(\frac{dp}{d\phi}\right)_v,$$

which give

$$\left(\frac{dt}{dv}\right)=-\left(\frac{dp}{d\phi}\right), \quad \ldots\ldots\ldots (5)$$

where we may dispense with the suffixes. If the right-hand side of this equation be simultaneously multiplied and divided by t, the denominator represents the amount of heat which is supplied, at constant volume, in order to produce the variation, dp, of pressure. If dp and $td\phi$ are positive, dt/dv is essentially negative. Hence the equation asserts that *substances which, at constant volume, have their pressure raised (or diminished) by the application of heat, will fall (or rise) in temperature during adiabatic expansion; and the change of temperature, per unit change of volume, is numerically equal to the product of the absolute temperature into the change of pressure per unit of heat supplied.*

We may now combine with equation (1) the equation

$$d(pv)=pdv+vdp,$$

as the result of which we get

$$d(e+pv)=td\phi+vdp \quad \ldots\ldots\ldots (6).$$

From this we deduce, as above, the result

$$\left(\frac{dt}{dp}\right)=\left(\frac{dv}{d\phi}\right) \quad \ldots\ldots\ldots (7).$$

If dt and dp are both positive, dv and $d\phi$ are necessarily of the same sign. Hence, multiplying and dividing the right-hand side by t, we see that *substances which expand (or contract), under constant pressure, when heat is supplied to them, rise (or fall) in temperature when they are subjected to adiabatic compression; and the change of temperature, per unit increase of pressure, is equal to*

the product of the increase of volume at constant pressure, per unit of heat supplied, into the absolute temperature.

If we now combine with (1) the equation

$$d(t\phi) = td\phi + \phi dt,$$

we obtain

$$d(e - t\phi) = -\phi dt - pdv \quad \ldots\ldots\ldots (8).$$

Therefore

$$\left(\frac{d\phi}{dv}\right) = \left(\frac{dp}{dt}\right) \quad \ldots\ldots\ldots\ldots\ldots (9).$$

Multiplying each side of (9) by t we see that *substances, which absorb* (*or emit*) *heat when their volume increases isothermally, have their pressure, at constant volume, raised* (*or diminished*) *by increase of temperature; and the change of pressure, per unit rise of temperature, is equal to the quotient by the absolute temperature of the heat which is absorbed* (*or emitted*).

Let L represent latent heat, and let $v' - v$ represent the change of volume of unit mass of the substance when it changes its state. In this case (9) becomes

$$\frac{\mathrm{L}}{t(v' - v)} = \left(\frac{dp}{dt}\right)_s$$

from which we can calculate the change of the melting-point, or the boiling-point, which results from a given change of pressure.

Combining (8) with

$$d(pv) = pdv + vdp,$$

we find

$$d(e - t\phi + pv) = -\phi dt + vdp, \quad \ldots\ldots\ldots (10),$$

which leads to

$$-\left(\frac{d\phi}{dp}\right) = \left(\frac{dv}{dt}\right) \quad \ldots\ldots\ldots\ldots\ldots (11).$$

This tells us that *substances which expand* (*or contract*) *under constant pressure when their temperature is raised emit* (*or absorb*) *heat in order that their temperature may remain constant when the pressure is increased; and the heat which is evolved* (*or absorbed*) *per unit increase of pressure is equal to the continued product of the temperature, the volume, and the expansibility.* For the expansibility is $1/v \,.\, dv/dt$.

301. In equation (2) the quantity f represents the rate at which the energy increases, per unit increase of temperature, at constant volume. It therefore represents the specific heat at constant

volume, which we have already denoted by the symbol c. Hence we have—from (1)—

$$cdt = td\phi,$$

with the condition $dv = 0$.

These equations may be written in the form

$$cdt = t\left[\left(\frac{d\phi}{dt}\right)_p dt + \left(\frac{d\phi}{dp}\right)_t dp\right], \quad \ldots\ldots (12),$$

$$0 = \left(\frac{dv}{dt}\right)_p dt + \left(\frac{dv}{dp}\right)_t dp \quad \ldots\ldots\ldots (13).$$

But the quantity

$$t\left(\frac{d\phi}{dt}\right)_p$$

is evidently the specific heat at constant pressure, hitherto denoted by k. Hence (12) becomes

$$0 = (k-c)dt + t\left(\frac{d\phi}{dp}\right)_t dp \ldots\ldots\ldots\ldots (14).$$

Now the condition $dv = 0$ necessitates a certain relation between dp and dt; but we can eliminate these quantities from (13) and (14).

Thus $$k - c = -\frac{\left(\frac{d\phi}{dp}\right)_t}{\left(\frac{dv}{dp}\right)_t}\left(\frac{dv}{dt}\right)_p.$$

By (11) this becomes

$$k - c = -t\frac{\left(\frac{dv}{dt}\right)_p^2}{\left(\frac{dv}{dp}\right)_t} \ldots\ldots\ldots\ldots\ldots (15).$$

To apply this result to the case of a perfect gas we must find the values of

$$\left(\frac{dv}{dt}\right)_p \text{ and } \left(\frac{dv}{dp}\right)_t$$

from the equation $pv = \mathrm{R}t$.

This gives $$\left(\frac{dv}{dt}\right)_p = \frac{\mathrm{R}}{p}, \text{ and } \left(\frac{dv}{dp}\right)_t = -\frac{\mathrm{R}t}{p^2},$$

whence $$k - c = \mathrm{R}. \ldots\ldots\ldots\ldots\ldots (16).$$

The difference between the two specific heats is therefore constant; and, since the values of both k and R can readily be found by experiment, c (the experimental determination of which is very difficult) can be calculated by means of this relation. (See § 271.)

302. The equation connecting the pressure, volume, and temperature of a perfect gas is $pv = Rt$. It will be useful to determine the relation between the pressure, the volume, and the entropy of such a gas. We have

$$d\phi = \left(\frac{d\phi}{dt}\right)dt + \left(\frac{d\phi}{dp}\right)dp,$$

where $t\,d\phi/dt$ is obviously the specific heat at constant pressure, and, by (11), $-d\phi/dp$ is equal to dv/dt at constant pressure, which again is equal to R/p, *i.e.*, to $(k-c)/p$. Hence

$$d\phi = k\frac{dt}{t} - (k-c)\frac{dp}{p},$$

the integral of which (§ 38) is

$$\phi = -\log a + k \log t + (k-c) \log p,$$

where a (and therefore $\log a$) is a constant. We may write this in the form

$$t^k p^{-k+c} = a\varepsilon^{\phi}.$$

But t is equal to pv/R, whence

$$p^c v^k = A\varepsilon^{\phi},$$

A being equal to aR^k. Thus, instead of $pv =$ constant, we must write

$$pv^{\frac{k}{c}} = \text{constant}$$

when adiabatic compression or expansion takes place.

303. When air is compressed by the sudden application of pressure, the heat developed is almost precisely equivalent to the work which is spent in producing the compression. Joule proved this by enclosing air in a strong vessel which could be placed in communication with another vessel, of the same size, which had been exhausted of air. Both vessels were placed in a large mass of water the temperature of which was accurately determined. When a stopcock in a tube connecting the two vessels was opened, the air rushed from the one vessel into the other so as to equalise the pressure throughout. The temperature of the vessel containing the expanding air was lowered, for work had been done during expansion, so that the air was cooled. That of the other

vessel rose, for the violent impact of the air which rushed into it caused the development of heat. But the amount of heat which was absorbed in the one case was almost precisely equal to that which was evolved in the other, for the surrounding water, which was well stirred, showed no appreciable change of temperature.

A more accurate form of this experiment was subsequently adopted by Joule and Thomson in their researches on the thermodynamical properties of gases. The gas under investigation was made to pass very slowly through a tube, in which a plug of cotton wool was placed, and its pressure and temperature on both sides of the plug were observed.

The preceding methods lead to a simple equation connecting the changes of temperature and pressure with the volume, the absolute temperature, and the expansibility of the substance. The expansibility may then be expressed, by means of Charles' Law, in terms of the temperature on the Centigrade scale, if the range of temperature be so small that the Centigrade and the absolute degrees are practically equal throughout its extent; so that the equation gives a direct comparison of the absolute and the Centigrade scales.

Boyles' and Charles' Laws give (§ 266) for a perfect gas $T = t + 1/\alpha$, where T is absolute temperature. The investigation just alluded to gives

$$T = t + \frac{1}{\alpha} - \psi$$

where ψ, as Thomson and Joule's experiments indicate, is, in true gases, a small quantity—generally positive.

The experiments showed that all the true gases except hydrogen were made colder by their passage through the plug, and indicated that the absolute zero is about 273°·7 C.

304. The truth of the Second Law of Thermodynamics (§ 292) rests entirely on the immensity of the number of particles contained in any portion of matter which is of a size comparable with the dimensions of our instruments and machines. An ordinary thermometer, placed in any position in a mass of air, might indicate uniformity of temperature; while another thermometer, sufficiently small in size, might (rather, *would*) indicate rapid variations of temperature, and might even show that heat was passing from cold parts to hot parts of the given mass. For, the quickly moving molecules might occupy on the whole one portion of a volume so small as to contain only a few molecules, while the slowly moving molecules occupied the remainder; and some of the slowest of the

quickly moving molecules might be exchanged for such of the slowly moving molecules as were actually moving more quickly than they were.

An *average* uniformity is preserved on the large scale, though, on a sufficiently small scale, it does not obtain. It is because of this average uniformity that the statement—that a heat-engine cannot continually draw the heat which it transforms into work from a body colder than its condenser—which Thomson made the basis of the Second Law of Thermodynamics is true.

CHAPTER XXVII.

ELECTROSTATICS.

305. *Electrification by Friction.*—When a rod of glass is rubbed with flannel—or, better, with leather coated with a paste of zinc amalgam—it acquires the property of attracting surrounding bodies. Light bodies, such as pieces of paper, can even be raised up by it against the attraction of the earth. When in this state, the glass is said to be *electrified*—or it is said that *electricity* has been developed upon the glass.

If the glass had been rubbed with cat's-skin or with any one of several other substances, similar effects would have ensued; but the extent to which electrification is developed depends upon the nature of the substance which is used as a rubber.

The glass may be replaced by sealing-wax, resin, ebonite, etc., and the phenomena will still be exhibited to a greater or less extent.

In all cases it is necessary for success that the substances shall be well warmed and dried.

306. *Conductors and Non-Conductors.*—If, instead of a glass rod or a rod of sealing-wax, we take a metallic rod, no electrical effects are in general observable; and many other substances also are incapable (unless special means are adopted, § 324) of being electrified by friction.

We are thus led to divide all substances into two classes according as they are or are not electrifiable by friction in the usual way. Those of the former class are called *Insulators*, *Dielectrics*, or *Non-conductors;* those of the latter class are called *Conductors*. The latter terms are applied because it is found that all substances which cannot usually be electrified by friction have the power of allowing electricity to flow along them, while the other class of substances prevent such flow.

307. *Fundamental Phenomena presented by Electrified Bodies.*—The substances which are attracted by an electrified body are not

necessarily non-conductors. In order to investigate the subject further we shall suppose that a pith-ball (which is a conductor, and is at the same time very light, so that the effects to be observed are easily seen) is the body to be attracted, and we shall suppose it to be insulated by being suspended from a dry glass rod by means of a dry silk thread.

If an electrified glass rod be brought into the neighbourhood of the pith-ball, the ball will be drawn towards it; and this will also take place when electrified sealing-wax is presented.

Now let the glass rod be brought so near that the ball comes in contact with it. Immediately after contact the ball is violently repelled by the glass; but, if the ball be touched with the hand, attraction will again occur; and the same phenomena will happen when electrified sealing-wax, or any other electrified body, is used. Still, though all electrified bodies produce this effect, a slight modification of the experiment will bring to view a profound difference in the nature of the electrification of different substances.

Instead of touching the pith-ball when it is repelled by the glass, let the electrified sealing-wax be presented to it. Strong attraction becomes apparent. Similarly the electrified glass will attract the ball when the sealing-wax repels it. And all substances which can be electrified by friction can be classified according as they act in this respect like glass or like sealing-wax.

308. *Positive and Negative Electricity.*—In order to explain these phenomena we make the following assumptions: 1st. There are two 'kinds' of electricity; 2nd. Like kinds repel each other, unlike kinds attract each other; 3rd. The attraction and repulsion diminish as the distance increases; 4th. An unelectrified body may be looked upon as a body which contains equal amounts of both kinds of electricity, which can be separated, to a greater or less extent, by means of the action of electrified bodies.

Let us distinguish the electricities developed on glass and sealing-wax as *positive* and *negative* respectively. When the positively electrified glass rod is brought near to the unelectrified pith-ball, we assert, in terms of our hypothesis, that the neutral electricities in the ball are separated, negative electricity coming to the side near the glass, positive electricity being repelled to the far side. The attraction between the unlike kinds is stronger than the repulsion between the like kinds, for the former are at a less distance apart. Hence the pith-ball moves towards the rod, for the electricity is confined to it and so cannot further alter its distance from the electricity of the rod unless the ball moves. If contact takes place, the negative electricity, which has been *induced* (as the phrase is) in

the ball, unites with a portion of the electricity of the rod, so that the ball is now charged with positive electricity, and is therefore repelled from the rod until it loses its charge (say, by repulsion to the ground when the observer touches the ball).

The same reasoning, with the interchange of the words positive and negative, applies when sealing-wax is used instead of glass.

Finally, the ball which has touched the glass rod is positively electrified, and is, therefore, attracted by the sealing-wax; and the ball which has touched the sealing-wax is negatively electrified, and so is attracted by the glass rod. All the phenomena are thus explained by means of our assumptions.

In the fourth assumption it was stated that an unelectrified body contains equal quantities of both kinds of electricity. In accordance with this assumption, it may be proved, by the methods to be shortly described, that the rubber which is used to produce electricity by friction becomes electrified to exactly the same extent as the rod which is rubbed, but with the opposite kind of electricity to that which is developed on the rod.

At one time it was customary to speak of electricity similar to that usually developed on glass as 'vitreous,' and of electricity similar to that which is produced on sealing-wax and other resins as 'resinous,' electricity. The mere fact that the so-called resinous electricity may be obtained from glass is sufficient proof of the undesirability of this classification. The terms *positive* and *negative*, as we have employed them above, are much preferable, for the words imply nothing but a distinction in kind.

The phrase 'kind of electricity' is very apt to be misleading. We do not yet know what electricity is. One would never dream of saying that the resultant positive and negative forces which constitute a stress are essentially different from each other, or that left-handed (positive) rotation is intrinsically different from right-handed (negative) rotation. Yet equal and opposite forces, and equal and opposite rotations, annul each other's effects. The terms 'positive electricity' and 'negative electricity' are merely adopted in order to enable us to consistently and concisely describe and (so far) explain certain phenomena.

The use of the old expression 'electric fluid' is to be carefully avoided.

309. *The Gold-leaf Electroscope.*—An *electroscope* is an instrument which is used to indicate the existence of electrification. If, in addition, the instrument measures the magnitude of the electrification, it is called an *electrometer*.

The gold-leaf electroscope is one of the most delicate of all electro-

scopes. It consists of two pieces of gold-leaf *a*, *a* (Fig. 166), which are connected, by means of a metal rod, to a metal head *h*. The rod passes through the top of a glass vessel in the manner indicated in the diagram. The glass vessel is open at the bottom, and contains a

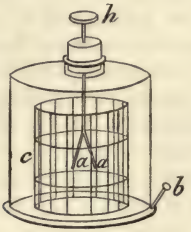


FIG. 166.

wire cage which surrounds the gold leaves. The cage can be placed in connection with the ground by means of the metallic connection *b*. The use of this cage will appear afterwards (§ 316); in the meantime we are merely concerned with the manner of using the instrument and the nature of its indications.

If a positively electrified body be brought into the neighbourhood of the head *h*, negative electricity is drawn towards the head, and positive electricity is repelled into the leaves, which diverge, since they are similarly electrified. The closer the body is brought to the head, the more widely do the leaves diverge; and, when the body is withdrawn, they collapse.

If, while the leaves are still diverging because of the presence of the electrified body, the head *h* be momentarily touched by the hand, instant collapse of the leaves will ensue (for the positive electricity escapes from the leaves through the hand to the ground), and the state of collapse will continue so long as the electrified body is not withdrawn. But when the body is withdrawn from the neighbourhood of the head, the leaves once more diverge; for the negative electricity which was drawn to the head of the instrument now spreads in part through the metal rod into the leaves. The latter therefore are diverging with negative electricity.

In this condition the instrument can be used to indicate the nature of the electrification of any body which is brought into the neighbourhood of the head *h*. If the body be negatively electrified, more negative electricity will be repelled into the leaves which will therefore diverge more. If it be positively electrified (or unelectrified), the negative electricity is drawn from the leaves which then

collapse; and if the positive electrification be sufficiently strong, some of the neutral electricity in the rod will be separated, and the leaves will diverge because of being positively electrified.

The interchange of the words positive and negative in the above reasoning will enable it to apply to the case in which a negatively electrified body is originally brought near to the head of the instrument.

It is obvious that these experiments are, in large part, merely a modified repetition of those which were discussed in § 307.

310. *Electrification by Contact and by Induction. Electric Quantity.*—In § 308 we have spoken of the electricity which is *induced* upon a conducting body because of the presence of another electrified body. So long as contact does not take place between the two bodies, the total amount of induced electrification is zero, a certain amount being drawn to one side of the body, while an equal amount of the opposite kind is repelled to the other side. But whenever contact occurs, the attracted electricity unites with some of the electricity in the inducing body; and so the conductor is electrified with the same kind of electricity as that which the inducing body possesses. It is then said to be electrified by contact. The total effect is the same as if the inducing body had given some of its electricity to the conductor, and it is usual to say that it *has* done so; for we cannot distinguish one amount of electricity from any other equal amount.

In the process of electrification by contact, the one body loses a certain amount of electricity, while the other gains an equal amount. This can be proved by means of measurements of the forces of attraction or repulsion which they exert upon an electrified body the electrification of which does not alter. In fact, we can electrically *weigh* out equal amounts of electricity, just as we can gravitationally weigh out equal amounts of matter. We are therefore justified in speaking of electricity as a thing which can be doled out in measurable quantities; and it is usual to say that an electrified body is *charged* with electricity, and to call the total quantity of electricity which it possesses its *charge*.

Suppose, now, that we have a charged body—charged positively, let us say. It is possible by its means to charge other bodies, either positively or negatively, to any desired extent—and that without any reduction of its own charge.

Let A (Fig. 167) be the positively charged body, and let B and C represent two of the other bodies, of which B is to be negatively charged, while C is to be positively charged. Each of the three bodies being well insulated from other conductors, place B and C in

contact in some such position relatively to A as is indicated in the figure. Then let B and C be separated: B will be charged negatively, while C will have a positive charge. Greater effects would be produced, if necessary, by placing B and C at a considerable distance apart, and joining them by a thin conducting wire; for the effect of the charge in A is largely counteracted by

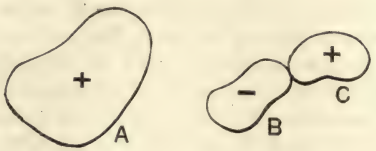


FIG. 167.

the mutual attraction between the positive and negative electricities in B and C; and this mutual attraction is diminished as the distance between B and C increases. In practice, it is convenient to let C be the earth, and to place B in connection with it by means of a metallic wire or other conductor. In this case the repelled positive electricity in C is practically at an infinite distance. We may then use the body B, instead of A, if we wish to charge any other body negatively.

This process, in which the inducing body does not lose any of its charge, is called charging by *induction*.

The induced charge is, except in one special case (§ 316), less than the inducing charge.

311. *Continued Production of Electricity. The Electrophorus.* —In last section we saw how it is possible to obtain a positive or a negative charge at will by means of a single insulated charged

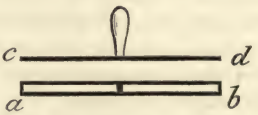


FIG. 168.

body and two conductors. The instrument, based on this principle, which is generally used for the purpose, is the *electrophorus*. It consists of a flat, circular, cake of resin, contained in a shallow metal vessel *ab* (Fig. 169). The resin is slightly warmed, and is electrified negatively by friction with cat's skin. A metal disc *cd*, insulated by means of a glass handle, is used instead of the con-

ductor B of last section, while the earth takes the place of the conductor C.

As the metal disc is brought near to the electrified resin, positive electricity is induced on its near side and negative electricity is repelled to its far side; and the more closely the disc is approached to the resin, the greater is the resultant separation of electricity.

When the disc is laid upon the resin, contact is made between it and the ground by means of a metal pin which passes through the centre of the cake of resin and is connected with the metal vessel enclosing it, and therefore with the ground upon which the vessel rests. The negative electricity escapes to the ground, and the disc is left charged with positive electricity, the charge being practically equal to that on the resin.

The disc may now be lifted away, and positive electricity can be communicated by contact from it to any conductor; and the process may be repeated from the beginning.

Two points in this explanation may present some difficulty. It may appear that the negative electricity of the resin should be destroyed whenever the disc is placed upon the surface of the cake; for the induced positive electricity in the disc would combine with it, and then the remaining negative charge in the disc would pass to the ground through the metallic connecter. This would really happen if the two surfaces came in contact throughout their whole extent; but, because of inequalities, they only touch over a comparatively small area. Again, the negative electricity in the disc is repelled by the electricity on the resin. How, then, can it pass to the ground by a connection which passes through the resin? We cannot be content with the reply that the disc and the earth are then both parts of one conductor, so that a road is opened up by which the electricity can get to a greater distance from that which repels it. This statement seems very like a statement to the effect that a stone would roll a short distance up one side of a hill, in order that it might get a longer roll down the other side. The fact that electricity flows like an incompressible fluid (§ 335) makes an explanation easy. There are three parallel layers of electricity in the apparatus—two negative layers with one intermediate positive layer. The lower negative layer tends to draw positive electricity from the ground; the intermediate positive layer repels it; and we are left with the upper negative layer which attracts it. We may suppose, therefore, that the negative layer on the upper side of the disc draws positive electricity from the ground and unites with it. The earth is consequently left with an equal negative charge, and the effect cannot be distinguished from

that which would have ensued upon an actual passage of the negative electricity of the disc to the ground. We have no means of distinguishing between the two cases, and therefore we are justified in saying that the electricity does pass from the disc to the ground.

Returning from this digression, we remark that the production of electricity by means of the electrophorus becomes more and more continuous the more rapidly the various motions of the disc are performed. The principle of all machines used for the production of a statical charge is the same as that of the electrophorus, but they are so constructed as to give a strictly continuous production.

312. *Law of Electric Attraction and Repulsion.*—We have hitherto explained the various facts which have come under our consideration by the assumption of attractive or repulsive force, which diminishes in intensity as the distance between the attracting or repelling quantities increases; and we found that the assumption enabled us to give a consistent account of the facts. Therefore, adopting this assumption as a working hypothesis, we must now consider more minutely the exact law of force.

The law was elaborately investigated by Coulomb by means of his torsion balance. In this instrument a vertical wire attached to the

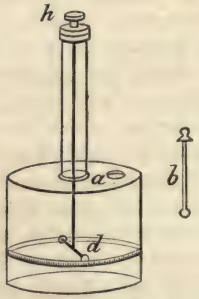


FIG. 169.

torsion-head h (Fig. 169) carries a horizontal insulating arm, at the end of which a small metal disc d is fastened.

A scale fastened to the glass cover which surrounds the instrument enables us to determine the angular position of the arm; and the position of the torsion-head, when there is no torsion on the wire, is also noted. Now let a positive charge be given to the disc d, and let another positive charge, contained in a metal ball fixed to

an insulating handle b, be introduced into the interior through the aperture a in the glass cover of the instrument, the length of the handle being such that the ball and the disc are in one horizontal plane. The mutual repulsion of the two quantities of electricity will cause the arm to twist round through a certain angle. Additional torsion is then put on the wire, by turning the head round, until the disc is brought back to its former position.

Now increase the charge in the ball in any ratio and repeat the same series of operations, having previously turned back the torsion-head into its old position. It will be found that the torsion, which must now be put on the wire in order to turn the disc back to its first position, is increased in the same ratio. This proves that the force is proportional to the quantity of electricity in the ball, and therefore, also, that it is proportional to the quantity in the disc.

Next, perform a series of experiments in which the charges of both bodies are kept constant, while their mutual distance is varied. The amount of torsion which is requisite at the different distances will show that the force varies inversely as the square of the distance.

The same law will be found to hold when the charges are negative, and also when one is negative and the other positive. Of course, in the latter case, the angular rotations of the arm and the torsion-head are necessarily reversed in direction.

Let q, q', be the quantities, and let s be the distance between them. The law of force is expressed by

$$+\frac{qq'}{s^2},$$

where the positive sign corresponds to repulsion and so indicates that the force is attractive (*i.e.*, is in the direction of *decreasing* distance) when q and q' are of opposite sign, and that it is repulsive (*i.e.*, is positive outwards) when they are of like sign.

This is the well-known law of gravitational force, and therefore all results which we have deduced (Chap. VIII.) regarding that force will at once apply to the case of electric force, provided that we take account of a possible reversal of sign.

313. *Electric Potential. Electromotive Force.*—If we attempt to increase the charge of an insulated conductor by any stated means in which the same conditions are maintained, as in the electrophorus, we find that it is more and more difficult to do so the farther the process is carried out, and that the charge cannot be increased beyond a certain limit. To make the reason for this clear we must make an apparent digression.

Electrified systems obviously possess energy in virtue of their electrification, for the mutual attraction or repulsion between their various parts may be used for the production of mechanical work.

We define the *Mutual Potential Energy* of two systems as *the amount of work which may be obtained from their mutual repulsion until they are at an infinite distance apart;* and we define the *Potential* at any point, due to a given electrical system, as *the mutual potential energy between the system and unit quantity of positive electricity placed at that point.* This definition makes the sign of the potential coincide with the sign of the electrification of the system to which it is due, and it makes the potential represent potential energy and not exhaustion of potential energy, as in the case of gravitation (§ 95).

Let V and V′ be the potentials at two points which are at a distance s apart. The average force which acts so as to transfer the unit of positive electricity from the point which is at potential V′ to the point which is at potential V is $(V' - V)/s$; and, if dV is the change of potential in the small distance ds measured from any point, the actual force at that point is

$$-\frac{dV}{ds}$$

for our definition makes V decrease as s increases.

This quantity is the rate of variation of potential per unit length, and is called the *Electromotive Force* at the given point, for it is the force which acts so as to transfer electricity.

We see therefore that no transference of electricity can occur between two conductors which are at the same potential.

Now we have seen that two like quantities of electricity, q and q', situated at a distance s apart, repel each other with a force

$$\frac{q'q}{s^2},$$

and the force with which q acts on a unit of electricity of like sign is

$$-\frac{dV}{ds} = \frac{q}{s^2}.$$

This gives

$$V = \int_s^\infty \frac{q}{s^2} ds = \frac{q}{s},$$

(see § 35).

Now, the potential of a conductor must be uniform throughout if the electricity which it contains is at rest, for otherwise an electro-

motive force would act from one part of it to another. Hence we are justified in speaking of *the* potential of a conductor; and the above expression shows that the potential of a conductor becomes larger and larger in strict proportion to the quantity of electricity which it contains, being positive when the charge is positive, and negative when the charge is negative. And it follows that the work, which must be expended in order to bring up to a conductor a given charge, becomes greater and greater as the charge already contained in the conductor increases.

In particular, if the potential of the body which carries up the new charge (the disc of the electrophorus) does not exceed a certain fixed value, the potential of the conductor to which the charge is given cannot be made, by this process, to exceed that fixed limit. Therefore the *charge* of that conductor cannot be increased above a fixed limit, which is the statement made at the commencement of this section.

The potential is zero only at an infinite distance; but, in order to practically carry a charge to an infinite distance, it is necessary merely to connect the conductor, which contains it, to the ground. For the earth is a conductor which is practically at an infinite distance, and any ordinary charge which is communicated to it produces no sensible variation of its potential. (See next section.)

314. *Capacity. Condensers.*—The quantity of electricity which must be given to a conductor in order that its potential may be raised by unity is called the *Capacity* of that conductor.

It follows at once, from the result of last section, that the capacities of two spherical conductors are in strict proportion to their respective linear dimensions. For, since electrostatic force obeys the same law as gravitational force (§ 312)—from which (§ 88) we know that the action of each quantity is the same as if it were condensed at the centre so far as points outside the sphere are concerned—the potential of a sphere, of radius a, is

$$V = \frac{q}{a}.$$

Hence, when $V = 1$, $q = a$; that is, the capacity is measured by the radius.

The capacity of a compound conductor which consists of two concentric spherical conductors may be made extremely great.

Let A and A′ (Fig. 170) represent the two surfaces, and let their radii be respectively a and $a+\tau$. Let A be charged with a quantity q of positive electricity, and let A′ be charged with an equal quantity

of negative electricity. These two charges will spread uniformly over the respective surfaces (§ 316).

Now the potential at any point of A′, due to its own charge, is $-q/(a+\tau)$; and its potential, due to the charge of A, is $q/(a+\tau)$.

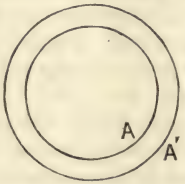


FIG. 170.

Its total potential is therefore zero; and so no flow of electricity will occur if it be connected to the ground by a conductor. This proves that if the sphere A be charged in any way, an equal and opposite charge will be induced on A′ if it be connected to the ground. Hence, if we add a positive unit of electricity to A, a negative unit will necessarily appear on A′ when it is 'put to earth'—to use the technical expression. But the work done in increasing the charge q of A by unity is q/a, and the work done in increasing the negative charge $-q$ of A′ by a negative unit is $-q/(a+\tau)$. Hence the whole work done is

$$w = \frac{q}{a} - \frac{q}{a+\tau} = q\,\frac{\tau}{a(a+\tau)}.$$

When τ is very small, this gives $q = a^2 w/\tau$; and, when w is unity, q represents the capacity of the arrangement, which is therefore

$$\frac{a^2}{\tau},$$

a quantity which may be made extremely great by sufficiently decreasing τ.

Any arrangement of this sort is called a *Condenser*, since it enables us, without much expenditure of work, to store up a large quantity of electricity. The only essential feature in any such arrangement is that the two conducting, and mutually insulated, surfaces shall be extremely close together in comparison with their own dimensions.

A common form of the condenser is known as the Leyden jar. This consists (Fig. 171) of a thin glass jar coated externally and internally with tin-foil. The neck and upper portion of the jar are not coated

with the foil, so that the insulation between the two conducting sheets may be as complete as possible. The mouth of the jar is usually closed with a cork, through which passes a metallic rod, terminating externally in a knob, and making communication in-

FIG. 171.

ternally with the inner coating of the jar. The external coating can be readily connected to the ground, while the internal coating is charged, through the agency of the rod, by means of any electric machine.

Discharge of the jar is effected by means of the *discharging-rod*, (Fig. 172), which consists of two jointed metal rods connected to a glass handle. The two knobs *a* and *b* are placed in connection with the outer and inner coatings of the jar; and the discharge, resulting

FIG. 172.

in the combination of the two equal and opposite quantities of electricity, takes place through the metallic circuit *acb*. Great care must be taken in the use of jars charged to a high potential, as very serious, if not fatal, effects might ensue upon their discharge through the human body.

Another common form of condenser consists of a pile of sheets of tin-foil separated by paper soaked in paraffin. All the odd sheets in the pile (counting from one end) are connected together. So also are the even sheets, but the odd and the even elements are carefully kept separate. This arrangement constitutes a condenser of great capacity.

A practically identical arrangement may be made by joining together all the internal coatings and, independently, all the external coatings of a number of Leyden jars. The capacity of the whole is equal to the sum of the capacities of the separate jars.

If the jars be connected together in series—that is, with the outer coating of one joined to the inner coating of the next in order, and so on—the resultant capacity is only equal to the capacity of each jar (all being supposed to be equal in this respect), and the whole charge is equal to the charge of one jar when its coatings are raised to the same (total) difference of potential. When the capacities are not alike, we may let V_0 and V_1 indicate the potentials of the outside and inside coatings of the first jar, whose capacity in C_1, while V_1 and V_2 are the similar quantities for the second jar, the capacity of which is C_2, and so on, V_n being the potential of the inner coating of the last jar. If C be the total capacity, $C(V_n - V_0)$ is the total charge. But this is equal to the sum of the separate charges. Hence

$$C(V_n - V_0) = C_1(V_1 - V_0) + C_2(V_2 - V_1) + \ldots\ldots\ldots + C_n(V_n - V_{n-1}).$$

But the charge of any one jar is necessarily equal to that of any other; for the outside coating of each has a charge which is equal and opposite to that of the inner coating of the jar to which it is joined, since the two form a single insulated conductor; and the charges in the two coatings of any one jar are also necessarily equal. Hence we have the $n-1$ equations

$$C_1(V_1 - V_0) = C_2(V_2 - V_1) = \ldots\ldots\ldots = C_n(V_n - V_{n-1}).$$

In all there are n equations connecting the $2n+1$ quantities C, $C_1, \ldots, C_n, V_0, \ldots, V_n$. If $V_0 = 0$, while the quantities $C_1 \ldots C_n$, are known, we can eliminate the Vs and calculate C.

315. *Specific Inductive Capacity.*—In last section we obtained the expression a^2/τ for the capacity of an arrangement consisting of two concentric conducting spherical surfaces of mean radius a and separated by an insulating interval of thickness τ.

It is customary to consider air as the standard insulating material.

If the air be replaced by some other insulating material, such as glass, resin, etc., it is found that the capacity of the same arrangement becomes

$$C = K\frac{a^2}{\tau}$$

where K is a constant for that particular medium and is called its *Specific Inductive Capacity.*

The Specific Inductive Capacity of a substance may therefore be determined by means of measurements of the quantity of electricity which is required to charge a jar, of which the given substance forms the insulating medium, up to a given potential. The ratio of this quantity to the quantity which is required in order to raise the potential of a precisely similar jar to the same extent, when air is the insulator, is the value of the required constant.

Faraday determined the Specific Inductive Capacity of various substances by measurements of potential. He used two precisely similar and equal Leyden jars which were so constructed that the insulating medium could be changed when desired. The coatings of one of the two jars were insulated by air; those of the other were insulated by a substance whose (unknown) inductive capacity K, was to be found. He charged the air jar to potential V and then divided its charge between itself and the other jar by making connection between their outer coatings and their inner coatings respectively. V′ being the resultant common potential of the jars, while C is the electrostatic capacity of the air jar, the equation

$$V'C+V'KC=VC$$

expresses the condition that the total quantity of electricity in the two jars is equal to that originally possessed by the air jar. Hence

$$K=\frac{V-V'}{V'}$$

is found in terms of the known potentials.

The value of this constant (sometimes called the Dielectric Constant) might also be found by observations of the reduction of the potential of a condensing arrangement, with a given charge, when the layer of air between the oppositely electrified surfaces is partially displaced by a layer of another insulating material. If t be the thickness of the new layer, the effective thickness of air displaced is t/K.

If vacuum be taken as the standard insulating medium, the specific inductive capacity of air is 1·00059, according to the determinations of Boltzmann. Faraday, with the apparatus at his disposal, was unable to observe any difference between the inductive capacities of different gases; but Boltzmann has shown that small differences really occur.

The dielectric constants of solids and liquids are greater than those of gases and differ considerably among themselves.

316. *Distribution of Electricity on Conductors. Electric Density.*—A static charge of electricity, which is communicated to

a conductor, is necessarily confined to the surface of that body. For, otherwise, the mutual repulsion between like quantities of electricity would produce continuous currents of electricity in the interior of the conductor.

The fact that the electricity is at rest on the surface of the conductor also shows that the distribution must be such as to produce a uniform potential all over the conductor.

If the conductor be spherical, we see, from the principle of symmetry, that the surface distribution of electricity is uniform; but if the conductor be not symmetrical, we would infer that the quantity of electricity distributed, per unit of area, over the surface cannot be uniform. This quantity is called the *density* of the surface distribution. It is large where the curvature of the surface is large, and is small where the curvature is small.

In order to determine the density at various parts of any surface (whether conducting or not), we might place a small plane metallic disc, attached to an insulating handle, in contact with the surface. If the curvature of the surface be not too great, and if the disc be sufficiently thin and small, the electricity on the part of the surface covered by the disc will be transferred to the outside surface of the disc, for it is then practically a portion of the electrified body. The disc may then be removed, and the magnitude of its charge may be tested. If the charge be q, while the area is a, the density at that part of the surface is q/a. The process may be repeated as often as we please, at different portions of the surface, so long as the total charge of the conductor is not sensibly diminished; and this objection might be entirely avoided by taking precautions to have the potential of the conductor maintained constant. As this method was first used by Coulomb, the disc is called Coulomb's Proof Plane.

In a number of cases the law of distribution of density can be calculated from the known electrostatic laws. Some of these cases will be discussed in the next two sections.

In particular, it is known experimentally that a closed conductor, which has an insulated charge placed inside it, and is connected to the ground, will become charged in such a way that no force is exerted at any external point by the internal and superficial charges; and the reason is that its potential is zero, so that the potential is uniformly zero in all external space. Hence the surface distribution, with its sign changed, acts at all outside points precisely as the internal charge does. Similarly the conductor screens internal bodies from the influence of external charges.

317. *Electric Images.*—We have already seen that a uniform

distribution of electricity over a spherical surface produces the same effect at external points as an equal quantity of electricity condensed at its centre would produce. This imaginary quantity is called the *electric image* of the uniform spherical distribution. [There is a reason for calling this quantity imaginary beyond the mere fact that it does not actually exist, and this is that the volume density of a finite quantity of electricity condensed at a point would necessarily be infinite, and such a condition cannot exist (§ 321).] We proceed now to the general discussion of the method of electric images, which is due to Sir W. Thomson, and which gives in many cases simple solutions of very formidable problems.

Draw a sphere (Fig. 173) with radius CM = a, and divide CM externally and internally in the points A, A′ repectively, so that

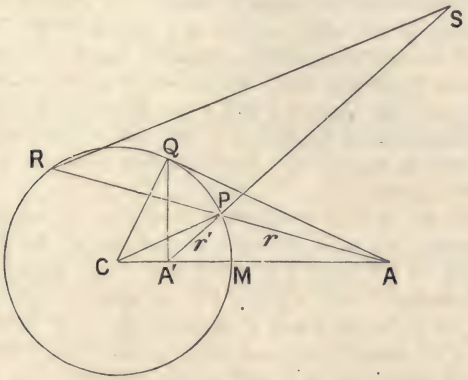


FIG. 173.

CA . CA′ = a^2. Take any point P in the circumference of the sphere, and join PA, PA′, PC. Let PA = r, PA′ = r', CA = d. We know by geometry that the triangles CA′P, CPA are similar.

Therefore $$\frac{\text{CA}'}{r'} = \frac{a}{r}.$$

Now put $$\frac{e'}{r'} = -\frac{e}{r},$$

and we get $$e' = -e\frac{\text{CA}'}{a}.$$

But if we draw A′Q perpendicular to CA, we have QA perpendicular to CQ, and therefore $CA'/a = CA'/CQ = CQ/CA = a/d$. Hence

$$e' = -e\frac{a}{d}.$$

If a quantity e of electricity be placed at A, while a quantity e' is placed at A′, the condition $e'/r' + e/r = 0$ asserts that the potential at the point P, due to these charges, is zero; and therefore the sphere is a surface of zero potential.

Consequently, if the sphere be an infinitely thin conductor, and be connected to the ground, the action of the charge e placed at A will induce upon it a charge, the effect of which at external points will be the same as that of e' placed at A′; and the charge e' placed at A′ will induce on the sphere a charge, the effect of which at internal points is the same as that of e placed at A. And further, these two induced charges are precisely equal and similarly distributed, but of opposite sign; for, when the sphere is uncharged and insulated, and the charges e and e' are placed at A and A′ respectively, the potential of the sphere is zero, and no resultant charge is induced upon it, neither can electricity flow from any one part of it to any other part.

[To find the law of distribution of density which produces this effect, produce AP to R, describe a circle round the points R, A′, and A, and let A′P meet this circle in S. The triangles RPS and A′PA are similar. Also PR and PS are proportional to the forces, f and f', at P due to e and e' respectively. For these forces are e/r^2 and e'/r'^2 respectively, and $e/r = -e'/r'$. Hence the forces are inversely proportional to r and r', and therefore are directly proportional to PR and PS, by construction. Consequently, the resultant force F at P is proportional to SR, and so $SR/SP = F/f' = AA'/r$.

This gives $$F = \frac{AA'}{r}f' = \frac{AA'e^2}{e'}\frac{1}{r^3} = \frac{AA'e'^2}{e}\frac{1}{r'^3}.$$

Hence a sphere, the density of which is inversely as the cube of the distance from an external (or internal) point, attracts internal (or external) matter, as if it were condensed at that point.]

The point A′ is the *image* of A with respect to the given sphere, and by means of the relation $CA \cdot CA' = a^2$, we can find the image of any distribution of electricity.

For example, let there be a given distribution of electricity along the straight line $A_1 A_2$ (Fig. 174). The image of the point A_1 in a sphere, of radius a, the centre of which is at C, is A'_1, which is such

that $CA_1 . CA'_1 = a^2$. Similarly, $CA'_2 = A'_2 = a^2$, and so on. It is easy to prove that the line $A'_1A'_2$ is a portion of circle. But if

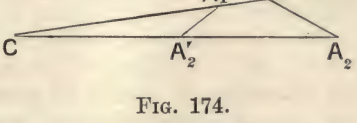


FIG. 174.

A_1A_2 is very small $A'_1A'_2$ is practically straight, and is inclined to A_1C at an angle equal to the angle of inclination between C_1A and A_1A_2.

Let $CA_1 = r$, $CA'_1 = r'$, $A_1A_2 = l$, $A'_1A'_2 = l'$, and we get, as the ratio of an infinitely small length in the image to an infinitely small length in the direct system,

$$\frac{l'}{l} = \frac{r'}{r} = \frac{a^2}{r^2} = \frac{r'^2}{a^2}.$$

Similarly, a' and a representing areas, v' and v representing volumes,

$$\frac{a'}{a} = \frac{a^4}{r^4} = \frac{r'^4}{a^4}; \quad \frac{v'}{v} = \frac{a^6}{r^6} = \frac{r'^6}{a^6}.$$

Next let λ, σ, ρ, represent line, surface, and volume densities in the direct system, and let λ', σ', ρ' be the corresponding densities in the image. We get

$$\frac{\lambda'}{\lambda} = \frac{\frac{e'}{l'}}{\frac{e}{l}} = \frac{e'}{e}\frac{l}{l'} = \frac{a}{r}\frac{r}{r'} = \frac{a}{r'} = \frac{r}{a}.$$

Similarly,
$$\frac{\sigma'}{\sigma} = \frac{e'}{e}\frac{a}{a'} = \frac{a}{r}\frac{r^4}{a^4} = \frac{r^3}{a^3} = \frac{a^3}{r'^3}; \quad \frac{\rho'}{\rho} = \frac{r^5}{a^5} = \frac{a^5}{r'^5}.$$

Again, if V and V′ represent direct and image potentials, we get

$$\frac{V'}{V} = \frac{\frac{e'}{r'}}{\frac{e}{r}} = \frac{e'}{e}\frac{r}{r'} = \frac{a}{r}\frac{r}{r'} = \frac{a}{r'} = \frac{r}{a}.$$

It is upon this relation, in terms of which *we can find the distribution of potential in the inverted system when we know the distribution of potential in the direct system*, that the physical use of the method of electrical images essentially depends.

For example, let a sphere be charged uniformly with electricity, and let us invert it with respect to itself. In that case, while the image coincides with the direct system, each point outside the image corresponds to a point inside the original sphere. But

$$V' = V\frac{a}{r'},$$

and we know that V is constant. Hence the potential at a point outside a uniformly charged sphere is inversely proportional to the distance of that point from its centre.

Next, invert the uniformly charged sphere with respect to an external point. The inverted system is also a sphere, and the density of the distribution upon it is given by

$$\sigma' = \sigma\frac{a^3}{r'^3}$$

where σ is constant. Hence the density of the image sphere varies inversely as the distance of the centre from the point of inversion. Further, points inside the first sphere invert into points inside the image. Hence the potential equation shows that the potential at an internal point varies inversely as the distance of that point from the centre of inversion, so that internal material points are attracted as if the sphere were condensed at the centre of inversion.

Finally, invert the uniformly charged sphere with respect to an internal point. The density of the image sphere varies inversely as the cube of the distance from the centre of inversion, and the potential varies inversely as that distance, but internal points have inverted into external points. Hence external material points are attracted by a sphere, whose density varies inversely as the cube of the distance from an internal point, as if the sphere were condensed at that point.

These three propositions have already been otherwise proved. A comparison of the present proofs with those previously given will exhibit to a slight extent the great power and simplicity of the method of electric images.

318. *Electric Lines of Force.*—As we have already seen, all the results obtained in Chap. VIII. regarding gravitational potential and force may be at once applied to the treatment of electrical potential and force. We may surround a given electrostatic system by equipotential surfaces, and we may suppose lines of force to be drawn, perpendicular to the equipotential surfaces, in such a way that the number of lines drawn outwards, per unit area, from such a surface represents the electric force at that part of the surface.

The results of that chapter then enable us to state that the density of electricity at any point of a charged surface is $1/4\pi$ times the number of lines of force which originate (or end) per unit area, on the surface at that point, the density being positive or negative according as the lines originate or end at the surface, that is, according as they are drawn outwards from, or inwards to, it. Similarly, the positive (or negative) volume density of electricity at any point of space is $1/4\pi$ times the number of lines which originate (or end) per unit of volume at that point.

Since no line of force can originate or end except at a point where electricity is situated, and since experiment shows that no electrical effect is felt outside a closed conductor, connected to the ground, which completely surrounds any electrical system, we see that all the lines of force which originate at the parts of the system must end (except in so far as they may proceed from one part of the system to another) upon the surrounding conductor; and, consequently, the charge induced on the conductor is equal in amount and opposite in sign to the total algebraic sum of the various internal charges. A particular case of this was discussed in § 314.

Every line proceeds from a point at high potential to a point at low potential. It therefore originates on a positively charged body and ends on a negatively charged body or passes to infinity.

319. *Electric Induction. Tubes of Induction.*—Hitherto we have spoken of gravitational and electrical action as if it took place directly at a distance. We have spoken of the mutual potential energy of two systems without inquiring *how* one system possesses energy in virtue of its position relatively to another. But, if we believe that energy is transferred by means of matter (§ 7), we must look upon some intervening medium as the vehicle through which it is transferred.

This is the way in which Faraday regarded the subject; and most of the development of electrical science in recent times is due to Faraday's work together with Clerk-Maxwell's mathematical interpretation and development of his views. The fact, stated in last section, that the electrical condition of charged conductors depends upon the nature of the intervening insulating medium gives strong support to this belief.

In Chap. XXIX. we shall see that, when a current of electricity flows along a conductor, the amount of electricity which crosses any section of the conductor, per unit of time, is constant. In other words, the flow of electricity resembles the flow of an incompressible fluid. Similarly, the facts that no quantity of one kind of electricity

can appear without the simultaneous development of an equal quantity of the opposite kind, and that the quantity of electricity induced on a closed conductor, which entirely surrounds the inducing electricity, is equal in quantity, and opposite in sign, to the latter, indicate that the induction of electricity through a dielectric resembles the displacement of an incompressible fluid. [This fact, probably, prolonged the use of the objectionable term 'electric fluid.']

We must therefore look upon a conductor as a body which cannot sustain electrostatic stress, and so permits electricity to flow along it when such stress is applied; and we are to regard an insulator as a body which can sustain electrostatic stress, in which 'displacement' of electricity takes place in proportion to the stress which is applied, and in which the displacement is annulled when the stress is removed. In this way of looking at the matter, a surface charge is supposed to reside on the surface of the dielectric and not on the surface of a conductor.

The word 'displacement' was introduced by Maxwell from the analogy to an elastic medium the parts of which suffer displacement when stress is applied, and recover from the distortion when the stress is removed. But Maxwell was very careful to avoid attaching any exact meaning to the term 'electric displacement.' He merely used it by analogy; and this cannot be too carefully kept in view.

Faraday used the term *Electric Induction* to indicate the state of the medium in virtue of which equal and opposite quantities of electricity appear on opposed surfaces; and Maxwell spoke of the total amount of induction through a given surface as the amount of electricity which is 'displaced' through it. A forward displacement in an insulator corresponds to a direct current in a conductor; a diminution of displacement corresponds to a reverse current.

The difference between the specific inductive capacities of various substances is explained by a difference in the amount of displacement which is produced in each under the same electromotive force.

We may draw lines of force through all points of any small closed curve on a conductor so as to form a tube of force; and we may draw such tubes, covering the whole surface of the conductor, in such a way that the number emanating per unit of area from all parts of the surface is equal to $4\pi\sigma$ when σ is the density of the electrification. The number of such tubes, which intersect unit area of any equipotential surface, therefore expresses the intensity of the force at that part of the equipotential surface.

But, instead of proceeding in the above manner, we may draw the tubes so that each encloses unit amount of electrification on the

conductor. Faraday called such tubes *Tubes of Induction;* for, when they originate, they enclose unit quantity of positive electricity, and, when they end, they enclose unit quantity of negative electricity. The total number of tubes of induction originating from, or ending on, a conductor, expresses its total positive, or negative, electrification, and the 'induction,' or 'displacement,' through each section of such a tube is constant.

Properly speaking, tubes of induction are formed by lines of induction, and not by lines of force. And it is well to remember that tubes of induction are not necessarily tubes of force: for the displacement does not always take place in the direction of the electromotive force, although in general it does. (Compare the elastic properties of non-isotropic solids, § 245.)

320. *Electric Energy.*—In order to estimate the amount of energy which is associated with the charge of a conductor at a given potential, we have merely to calculate the work expended in charging it. Let Q be the charge, and let V be the potential. Let us suppose that the conductor is charged by successive infinitesimal instalments dq, and that the charge at any instant is q, while the capacity of the conductor is C. The potential at the given instant is therefore q/C. But the potential is the work which is required in order to bring up unit charge from an infinite distance, and give it to the conductor. Hence the amount of work which is necessary in order to increase the charge q by the amount dq is

$$q\frac{dq}{C};$$

and the total amount of work which must be expended in raising the charge from O to Q is

$$\int_0^Q \frac{q\,dq}{C} = \tfrac{1}{2}\frac{Q^2}{C} = \tfrac{1}{2}CV^2 = \tfrac{1}{2}QV.$$

Hence the whole energy is one half of the product of the charge into the potential; or one half of the product of the capacity into the square of the potential; or one half of the quotient of the square of the charge by the capacity.

Now let us look at the problem from the point of view of induction. Consider a positively electrified body inside a closed conductor. Draw unit tubes of induction from the body to the internal surface of the conductor, and describe equipotential surfaces corresponding to all potentials which differ from each other by unity, and are included between the potential V of the electrified body and

the potential V′ of the surrounding conductor. The number of cells into which the tubes of induction and the equipotential surfaces divide the volume between the two charged surfaces is equal to the product of $V - V'$ into the number of tubes, *i.e.*, into the charge of the body. Hence the number of cells into which the space is divided is double of the electrical energy of the system.

A simple extension of this reasoning shows that the same result is true whatever be the number of electrified bodies contained inside the conductor. (See Maxwell's *Elementary Treatise on Electricity*, Chap. V.)

This result points to the conclusion that the energy of a system of charged conductors is contained, not in the conductors themselves, but in the insulating medium which surrounds them. And Faraday's and Maxwell's views of the nature of induction show us how this may be. The dielectric is in a state of strain so long as displacement is maintained by the action of electromotive force; so that the energy which was expended in producing the strain is contained in the dielectric in a potential form. [Visible strain may be produced in a piece of glass by means of electrostatic stress.]

To obtain an expression for the amount of energy contained, per unit of volume, in the dielectric, let us consider, as the simplest case, an insulated sphere of variable radius r charged with a constant quantity q of positive electricity. The potential of the sphere is q/r, and the energy contained in the space external to it is

$$\tfrac{1}{2}q \cdot \frac{q}{r}.$$

Now let r increase infinitesimally to $r+dr$. The energy becomes

$$\tfrac{1}{2}q \cdot \frac{q}{r+dr}.$$

The difference of these quantities,

$$\tfrac{1}{2}q^2 \frac{dr}{r^2},$$

is the energy which is contained in the intervening shell of volume $4\pi r^2 dr$. Hence the energy contained per unit of volume at the distance r is

$$E = \frac{1}{8\pi}\frac{q^2}{r^4} = \frac{1}{8\pi}F^2$$

where F is the resultant force at the distance r due to the charge q.

If K be the specific inductive capacity of the medium we must write

$$E = \frac{K}{8\pi} F^2.$$

This result is quite general, F being the resultant force at any point due to the total electrification.

Maxwell has investigated the nature of the stress in the dielectric which would account for observed electric phenomena. He finds that the stress consists of a *tension* $KF^2/4\pi$ along the lines of force coupled with an equal *pressure* in all directions at right angles to the lines of force.

In particular, the tension at the charged surface of a conductor is $4\pi K\sigma^2$ in a direction perpendicular to the surface, where σ is the surface density of the electrification.

As an instructive example in connection with the above expression for the energy of an electrified system, we may estimate the energy of a charge Q, first when it is contained in a jar of capacity C, and, second, when it is divided between that jar and another jar of capacity C′. The original energy is

$$\tfrac{1}{2}\frac{Q^2}{C}.$$

After division, since the potentials of the two jars are equal, we have

$$\frac{Q_1}{C} = \frac{Q_2}{C'},$$

where Q_1 and Q_2 are the charges in the jars of capacities C and C′ respectively. Also

$$Q_1 + Q_2 = Q,$$

and hence

$$Q_1 = Q\frac{C}{C+C'},\quad Q_2 = Q\frac{C'}{C+C'}.$$

And the respective energies are

$$\tfrac{1}{2}\frac{Q_1^2}{C} = \tfrac{1}{2}Q^2\frac{C}{(C+C')^2},\quad \tfrac{1}{2}\frac{Q_2^2}{C'} = \tfrac{1}{2}Q^2\frac{C'}{(C+C')^2}.$$

Therefore, the total energy is

$$\tfrac{1}{2}\frac{Q^2}{C+C'},$$

which is always less than the original energy. The fraction $C'/(C+C')$ of the whole energy has been dissipated in the process

of division—usually taking the form of sound, light, and heat. [Compare the dissipation of energy, which takes place when a gas is allowed to expand without doing work, as in Joule's experiment (§ 303), or that which occurs when heat diffuses, so as to arrive at a lower temperature without the performance of work. In the present case no electricity is lost, but the potential is lowered.]

321. *Electric Absorption. Disruptive Discharge.*—If an elastic medium be distorted beyond its limits of perfect elasticity, the removal of the stress is not followed by complete recovery from strain; but if the distortion be not too great, complete recovery may take place after a sufficient time has elapsed. Conversely, a long-continued force may produce large distortion.

Analogous phenomena appear in dielectric media when subjected to electrostatic stress.

Thus, a Leyden jar, when charged to a certain potential, will gradually fall in potential, though it is well insulated. The result is the same as if its capacity gradually increased, or as if the specific inductive capacity gradually increased so that the same displacement was maintained by a smaller difference of potential. If the jar be now discharged, the quantity of electricity which is obtained is smaller than the original charge. This phenomenon is known as *Electric Absorption*, for the jar *appears* to have absorbed some of its charge.

A second (small) discharge may be obtained from the jar if it be left for some time. Subsequently a third may be obtained, and so on until the total discharge equals the original charge. These are called *Residual Discharges*. The apparently absorbed charge seems to leak out again, as if the medium gradually recovered from a temporary distortion.

It is not well to pursue these analogies too far. Maxwell has shown that apparent absorption will take place if the insulating medium is heterogeneous in the sense that it consists of parts, the specific inductive capacities of which differ from one another, or of parts which differ from each other in their insulating power. The insulating power of ordinary dielectrics, such as glass, gutta percha, etc., is not perfect; and so, if a composite dielectric consisted of alternate layers of incompletely and completely insulating materials, electric absorption would be manifested.

We know also that elastic solids are only capable of withstanding strain to a limited extent, and that they will be ruptured if too great stress be applied. Similarly all dielectrics will cease to insulate electricity if they are subjected to too great electrostatic stress. The state of strain in the insulating material of a Leyden jar

becomes greater and greater as the potential of the jar is raised higher; but, if the process be continued too far, the insulation breaks down, and the separated electricities recombine through the ruptured dielectric. This phenomenon is called the *Disruptive Discharge.*

A Leyden jar, through the substance of which the disruptive discharge has occurred, is useless for all subsequent electrical purposes, for the glass is in part shattered by the discharge. On the other hand, if air or any other fluid were used as the dielectric, the jar would insulate as completely as ever it did so long as too great stress were not again applied; for, although by the energy of the discharge the parts of the fluid medium would be violently disrupted, the insulation would be restored by an inflow of the surrounding medium.

The disruptive discharge is usually accompanied by the production of sound, light, heat, and mechanical effect, the total energy evolved being the exact equivalent of the original electrical energy.

Various forms of the disruptive discharge exist. The most ordinary form is called the *spark* discharge. When two oppositely charged surfaces are brought sufficiently near each other, the electrostatic stress in the medium increases to such an extent that the electricities combine because of rupture of the dielectric between the charged surfaces. A small streak of light is apparent where the discharge occurs, its form depending upon the thickness of the dielectric through which the discharge occurs. When the distance is great the streak of light (the spark) is very irregular and jagged in outline.

Feddersen found that the nature of the spark discharge depends upon the resistance (§ 335) of the circuit in which the discharge occurs. When the resistance is sufficiently large, it consists of successive rapid discharges in the same direction. It becomes continuous when the resistance is lessened to a certain extent, and it consists of a rapidly alternating series of discharges in opposite directions, when the resistance is still further diminished.

Sometimes the discharge is in the form of a *brush.* This is seen chiefly when one of the two conductors has great curvature at the place where the discharge occurs. A short line of light, which abruptly branches out into a brush-like form, appears at the place. Wheatstone showed, by means of his revolving mirror, that the brush discharge consists of a series of rapidly succeeding separate discharges. Its intermittent character gives rise to the crackling, or even musical, sound which accompanies this form of the discharge.

The *glow* discharge takes place from the rounded extremity of a wire which projects into the air. The end of the wire is covered by a phosphorescent light. This form of the discharge does not appear to be intermittent. It seems rather to be, as Faraday concluded, a convective discharge, in which the charge is carried away by the particles of the air. (Compare the action of the pith-ball, § 307.)

The 'electric wind,' which blows from a sharp electrified point, is due to the repulsion of air particles which have been electrified by contact with the point.

The limiting tension (§ 320) which the insulating medium can sustain without rupture is called the *Dielectric Strength* of the medium. The dielectric strength of air depends upon the distance between the oppositely electrified surfaces. It has a greater value when the distance is small than it has when the distance is large. In all gases it increases as the pressure increases, and diminishes as the pressure diminishes—but not indefinitely. A minimum value is reached at a certain stage, beyond which the strength increases as the pressure is farther diminished.

The method of spark discharge under diminished pressure (in so-called vacuum tubes) is much used for the purpose of examination of the spectra of gases.

322. *Atmospheric Electricity, etc.*—The atmosphere is almost always in a state of electrification, either positive or negative. The electrification is generally positive during long-continued fine weather; it generally becomes negative when the fine weather breaks.

In order to test the nature of the electrification, use may be made of Thomson's water-dropping accumulator. This instrument consists of an insulated metallic vessel, which contains water, and which is fitted with a long fine nozzle, from which the water issues drop by drop when the stopcock with which it is fitted is opened. The nozzle projects out into the external atmosphere by an opening in the window, and the vessel is connected with an electrometer. The stopcock is then opened and the water drops out.

If the atmosphere be positively electrified, negative electricity will be induced in the nozzle and, therefore, in the drop, while positive electricity is repelled to the electrometer. As each drop falls away, carrying its negative charge with it, the vessel and electrometer are left more and more positively charged. The electrification of the atmosphere is, therefore, indicated by the development of a charge of like sign in the electrometer.

An interesting question arises in this connection—What is the source of the energy of the charge in the electrometer? The energy of the charge may be transformed into heat, and the energy

of the falling drops may also be transformed into heat. Further, there is no other possible source of heat in the arrangement. But the drops may fall without any production of electric charge, and, therefore, the principle of conservation compels us to assert that the drops will fall more slowly when they are electrified than they do when unelectrified, and so will do less work. This conclusion is verified by experiment.

If we replace the metallic vessel (above alluded to) by a hot crucible, into which we drop water, the water will evaporate, and the vapour will be found to be negatively electrified, for the crucible and the electrometer become positively charged. If the vapour condenses, the total volume of all the drops of water which are formed remains constant; but the total surface of the drops diminishes as each drop increases in size, and the same quantity of electricity is confined to a smaller surface. The result is that the potential of the drops rises considerably. It is possible that the high potential of thunder-clouds may be explained in this way.

When the potential rises to such an extent that the air is unable to withstand the electrostatic stress, disruptive discharge (lightning) takes place.

The great use of a lightning-rod is to prevent the potential from rising to such an extent that disruptive discharge will occur. It does this by drawing off from the surrounding air a continuous current of electricity. The electrified air induces the opposite electrification in the rod, and the density is very great at the sharp point—so great, that the electricity streams off from it to the air by silent discharge, and so annuls, totally or partially, the electrification of the air; and this is equivalent to the passage of the opposite electricity from the air to the ground through the rod. If a cloud in the neighbourhood of the rod were suddenly electrified to a high potential by disruptive discharge from a distant thunder-cloud, the rod may not be able to draw off the electricity with a sufficient rapidity to prevent discharge from the near cloud to the building supposed to be protected by the rod. The rod would, more likely than not, be insufficient for the purpose of carrying off the discharge to the ground.

323. *Pyroelectricity.*—If a crystal of tourmaline, or of some other minerals, be heated, electrical phenomena will be manifested, although previously the crystal appeared to be unelectrified. Positive electricity appears at one end of the crystallographic axis, negative electricity appears at the other.

The electrification may now be destroyed by passing the crystal through a flame. Further electrification will then be manifested,

similar to that which formerly appeared, if the heating be proceeded with; but if, on the contrary, the crystal be allowed to cool, the opposite electrifications will appear at the ends.

Sir W. Thomson supposes that such crystals possess internal electrification—that they are electrically polarised in the direction of the axis—and that, when they are passed through the flame, their surfaces become electrified in such a way as to annul at all external points the effect of the internal electrification. And he further supposes that the amount of internal electrification depends upon the temperature, so that heating or cooling disturbs the balance of external effects.

324. *Electrification by Contact.*—The electrification of glass or of sealing-wax, etc., by friction may be explained by the assumption that an electromotive force exists at the surface of contact of the two substances which tends to produce electric displacement across the interface, and that friction is used merely for the purpose of securing better contact.

The substances being non-conductors, the electricity cannot pass from the surface, and the displacement continues until the effect of the reverse force which it entails balances the effect of the electromotive force of contact. So long as the surfaces remain in contact, the arrangement acts as a condenser of extremely large capacity, and relatively large displacement may be produced by a comparatively feeble difference of potential. But whenever the surfaces are separated, the potential rises greatly, because of the ensuing decrease of capacity. In this way the high potential obtainable from frictional machines, or from the electrophorus, etc., is explained. The charge is the same after separation as before it, but the potential has increased, and therefore the energy has increased; and the increase of energy is the precise equivalent of the work done in the process of separation.

Now, although we cannot electrify conductors by friction in the way that we electrify non-conductors, we can produce electrification of conductors by contact or friction, provided that we take proper precautions.

If we take two flat pieces of zinc and copper, insulate them both, and then place their flat faces in contact, the copper will become negatively electrified, while the zinc becomes positively electrified. This may be proved by separating the plates (still insulated) and testing their electrification by means of the electroscope or the electrometer. And we may explain the result by stating that an electromotive force of contact acts at the surface of separation of the metals in the direction from copper to zinc.

Volta found that this assumed electromotive force of contact between any pair of metals is equal to the sum of the electromotive forces between every pair of metals forming a series closed by the given pair. From this it would follow that the sum of the contact forces in any complete heterogeneous metallic circuit is zero. This is known to be true so long as the temperature is uniform throughout the circuit; and it is in accordance with the principle of conservation of energy, for there is no source of energy in such a circuit.

The assumption of the existence of an electromotive force of contact between metals sufficiently great to account for the observed effects is regarded as inadmissible by many physicists. That a contact force does exist is shown by thermoelectric phenomena; but this force is very much smaller than the Volta contact force. Consequently, those physicists who are unwilling to admit the possibility of a true contact force between metals, which would account for the whole observed effect, look upon the surfaces of separation between the metals and the air as the real seat of the electromotive force. The whole question is still involved in much uncertainty.

Contact forces exist between metals and liquids, and also between different liquids. Volta's law does not hold universally in the latter case.

325. *The Electrometer.*—Instruments such as the gold-leaf electroscope may be used for the purpose of obtaining very rough measurements of electromotive force. The *electrometer*, in one or other of its various forms, is used when accurate measurements of electrostatic effects are required. It is used directly for the determination of difference of potential and it may be indirectly used for the purpose of the comparison of the capacities of conductors, and consequently for the determination of their charges.

Instruments such as the gold-leaf electroscope are termed *idiostatic* instruments, for there is no electrification in any part of these instruments except such as is due to the electrification which is to be tested; and their indications (when small) are therefore proportional to the square of the difference of potential which is to be observed. Any small variation of potential is therefore inappreciable when the potential itself is small, and the indications of the instrument are the same kind, whether the potential is positive or negative. In *heterostatic* instruments, some of which we shall now describe, a constant charge of one definite kind is maintained in one part of the apparatus, so that a small variation of potential produces the same effect, whether the potential is large or small, and the indication is reversed in direction when the potential changes sign.

Most forms of the electrometer depend for their action upon the electrostatic force between similarly or oppositely charged bodies.

Coulomb's torsion balance is therefore one (but a very imperfect) form of electrometer.

In the *attracted disc* electrometer the two charged bodies are in the form of parallel horizontal discs placed at a distance apart which is small in comparison with their transverse dimensions. We shall assume that the discs are oppositely charged, the densities of the electrifications being $+\sigma$ and $-\sigma$ respectively. Except in the near neighbourhood of the edges, the lines of force are perpendicular to the discs, and the force at any point between them is (§ 99) $2\pi\sigma - 2\pi(-\sigma) = 4\pi\sigma$ in the direction from the positively charged disc to the negatively charged disc. Since $2\pi\sigma$ is the force with which the positively charged disc acts upon unit quantity of negative electrification on the other disc, the total force with which it attracts that disc is $2\pi\sigma.\sigma a = 2\pi\sigma^2 a$, where a is the area of the disc. But, as above, the force at any point between the discs is $4\pi\sigma$, and is equal to $(V - V')/t$, where V and V′ are respectively the potentials of the positively and the negatively charged discs, and t is the interval between them. Hence $\sigma = (V - V')/4\pi t$, and the total force of attraction between the discs is

$$F = \frac{1}{8\pi}\frac{(V-V')^2}{t^2}a,$$

which gives

$$V - V' = t\sqrt{\frac{8\pi F}{a}}.$$

[We might have deduced this result from the expression for the energy contained in unit volume of the dielectric (§ 320), which is

$$\frac{1}{8\pi}F^2 = \frac{1}{8\pi}\frac{(V-V')^2}{t^2}.$$

Hence the total energy contained in the volume at between the discs is

$$\frac{1}{8\pi}\cdot\frac{(V-V')^2}{t}\cdot a.$$

And the rate at which this varies per unit of thickness gives the force

$$F = \frac{1}{8\pi}\cdot\frac{(V-V')^2}{t^2}\cdot a.]$$

In Thomson's *absolute electrometer*, in which this principle is used, a concentric circular portion of the upper disc is alone moveable, and so the difficulty of the non-uniformity of the force at the edge of the

discs is avoided. The moveable part as nearly as possible fills the aperture without touching its sides, and it is suspended by means of a delicate (spring) balance, which has a fiducial mark by means of which the lower face of the movable disc can always be placed in one plane with the lower face of the surrounding portion of the upper disc—called the *guard-ring*. The lower disc can be accurately moved, perpendicular to its own plane, through known distances, by means of a screw. The balance and disc are surrounded by a metal case for the purpose of preventing any disturbance which might arise from external electrification. When the two discs are connected to bodies of different potential, the balance is depressed, and the screw is turned until it returns to its standard position. In this way the distance t is determined. Also, by previous experiments, it is known what weight must be placed on the disc in order to bring it to its standard position, and this gives the (constant) value of F.

In using the instrument it is preferable to keep its lower disc at a constant potential by means of a charged condenser—the constancy being determined by means of a secondary electrometer. The value of t is then found, first, when the upper disc is connected to the ground, and again, when it is connected to the body whose potential is to be determined. If θ be the difference of the distances, the above equation gives us the value of the potential

$$V = \theta \sqrt{\frac{8\pi F}{a}},$$

a being now the area of the attracted disc.

In Thomson's *quadrant electrometer* an aluminium needle swings inside a hollow metal cylinder, which is divided into four quadrants.

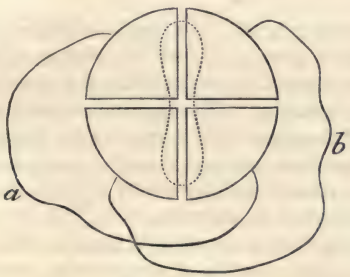


FIG. 175.

The two opposite pairs of quadrants (Fig. 175) are connected by wires. The needle in its normal position is suspended with

length directed along one of the lines of division between the quadrants, and it is charged to a high positive potential. One pair of quadrants, say those connected to the wire a, may be connected to the ground, while the other pair is connected by the wire b to a body at positive potential V. The quadrants connected with b become positively charged, and the quadrants connected with a become negatively charged. The needle is therefore deflected towards the negatively charged quadrants, and the deflecting couple is proportional to V, if the potential of the needle be sufficiently high.

Modifications of this instrument may be used for the measurement of extremely small, and of extremely large, differences of potential.

326. *Electric Machines.*—The electrophorus, which is the simplest form of electrical machine, has been already described.

As an example of the older machines, used for the continuous production of electricity, we shall take the *cylinder machine*. This machine consists of a glass cylinder C (Fig. 176), which is turned

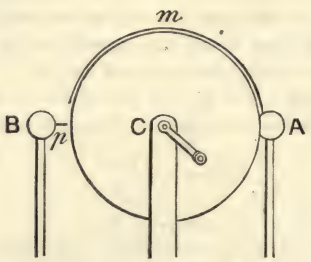


FIG. 176.

round in the direction AmB. An insulated leather rubber A, coated with zinc amalgam, presses against the rotating cylinder, and causes the development of positive electricity on the glass, while it becomes itself negatively electrified. The positive electricity is carried round on the surface of the glass until it reaches the sharp metal points p, which project from the insulated metallic conductor B. It induces in the points negative electricity which is discharged on the surface of the glass, destroying its electrification, and leaving B positively electrified. A silk flap m, connected to the rubber and resting in contact with the upper portion of the cylinder, prevents the electrification from slipping back along the surface of the glass. The potential of the positive electricity rises rapidly as it is carried from A to B, and the resulting electromotive force may cause the

electricity to slip back, so that the potential of B cannot rise very high. The silk flap becomes negatively electrified and so prevents the slipping, or stops it if it does occur by using the slipping electricity to destroy its own negative electrification, which tends to equalise the potential.

In the *plate machine* the glass cylinder is replaced by a circular glass plate, which is rubbed on both sides.

The *Holtz machine* is one of the best modern forms of electric machines. A fixed glass disc D (Fig. 177) has two paper armatures (a, a') fixed on it near the opposite extremities of a diameter. Near each armature an opening (shown by dotted lines) is cut in the glass, through which a paper point attached to the armature projects so as nearly to come in contact with a revolving glass disc C,

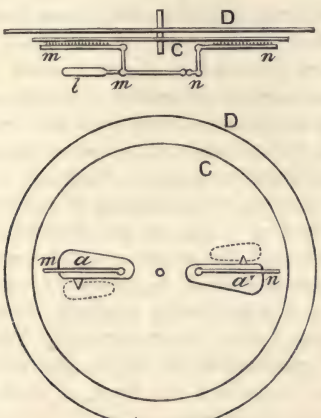


FIG. 177.

which is mounted upon an axis passing through the centre of D. A metal conductor m, fitted with a row of sharp points, faces the revolving disc on the other side opposite the armature a. A similar conductor n faces the armature a', and can be placed in communication with m by means of the rod l, which slides through the knob m.

In order to work the machine, the knobs n and m are placed in connection, and a charge (positive, say) is given to the armature a by means of the electrophorus or otherwise, while the disc is rotated in the direction aCa'. After a short time a rustling sound is heard, and the machine becomes difficult to drive. It is producing elec-

tricity, and the extra work which has to be performed is the equivalent of the electrical energy which is developed.

We may explain the process of charging in the following manner: The positive charge given to a induces negative electrification in the points of the conductor m. This is discharged upon the glass surface, and the glass carries it round to the opposite side of the machine, leaving m positively charged. Here it induces positive electricity in the sharp point of the armature a', and this is discharged upon the inside of the glass disc, leaving a' negatively charged. The armature a' now draws positive electricity to the points of the conductor n. This electricity is discharged upon the outer surface of the disc, leaving the conductor n negatively charged. Thus we may regard the glass disc as constantly carrying positive electricity from n to m in the one half of its revolution, and as constantly carrying negative electricity from m to n in the other half of its revolution.

It is possible that the negative charge is given to a' by way of the conductor n. The positive electricity, which is produced in m by the first motion of the machine, and flows to n, may be supposed to act inductively on the armature a', drawing negative electricity to the body of it, and repelling positive electricity to the sharp point to be discharged upon the glass, leaving a' negatively charged.

When the conductors n and m are slightly separated, a brush discharge passes through the air space. This brush discharge may be changed into a violent spark discharge by connecting the inner coatings of two Leyden jars to the conductors n and m, the outer coatings of the jars being joined together. The jars have to be charged up to the potential required for discharge through the air space, and, as their capacities are large, a great quantity of electricity passes at each discharge.

CHAPTER XXVIII.

THERMO-ELECTRICITY.

327. *Thermo-electric Phenomena.*—Though the principle of conservation of energy shows (§ 324) that the sum of the electromotive force in a closed metallic circuit must be zero, provided that there be no difference of temperature between the various parts of the circuit, we cannot assert that their sum will be zero if the temperature be not uniform. For, in the process of equalisation of temperature, it is possible that there may be transformation of thermal energy into electric energy; and this transformation will occur if the electromotive forces of contact between the metals which form the circuit depend upon the temperature.

Now, Seebeck discovered in 1822 that, in general, a current of electricity flows around a circuit, which is composed of two different metals, if there be a difference of temperature between the two junctions of the metals; and this shows that the equilibrium of the contact forces has been destroyed because of the variation of temperature.

And we cannot assert, without experimental evidence, that there will be no resultant electromotive force in a closed circuit composed of a single metal which varies in temperature from point to point. But the experiments of Magnus showed that no such force exists.

Still, in order that Magnus's result should hold, it is necessary that every part of the circuit should be *physically* similar and not merely chemically similar. For example, two portions of the same metallic substance, which are in a different state of strain, are physically different substances, and are also thermo-electrically distinct. And two portions of the same substance which are at finitely different temperatures are in different physical states, and might, therefore, exhibit thermo-electric phenomena. Such phenomena were observed by Le Roux and others at the instant when

contact was made between two portions of the same metal which differed abruptly in temperature.

328. *Laws of Thermo-electric Circuits.*—It is found by experiment that *the introduction of a piece of metal into a thermo-electric circuit does not contribute to the electromotive force of that circuit, provided that the extremities of the metal are at one and the same temperature.* We may, therefore, use solder to connect together the various parts of the circuit.

Let the lines A (Fig. 178) represent two pieces of the same metal. Let two of their ends have a common temperature t_1, while the other two have a common temperature t_o; and let the ends which are at the temperature t_1 be joined by a metal C, while the ends which are at the temperature t_o are joined by a metal B. As the whole arrangement is symmetrical with respect to the pieces A,

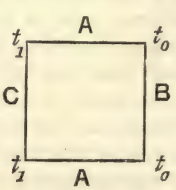


FIG. 178.

it is obvious that there can be no resultant electromotive force in the circuit. And if the temperature of one of the junctions between C and A be changed from t_1 to t_2, the metal B will still, from its symmetrical position, contribute nothing to the electromotive force, although there may now be a resultant electromotive force in the circuit.

Next let us arrange pieces of two metals alternately, as in Fig. 179, and let the temperatures of their extremities be as indicated. The pieces B, which are at temperatures t_1 and t_2 respectively, contribute nothing to the total effect, so that the whole arrangement really consists of two metals (A and B), the two junctions of which are t_3 and t_o respectively. Now we may join the pieces t_1Bt_1 and t_2Bt_2 respectively to those points of the piece t_3Bt_o, which are at the temperatures t_1 and t_2, by means of connections made of the metal B; and these new connections contribute nothing to the total electromotive force in the circuit. But the electromotive force in the part BBt_oAt_1B is due to the metals B and A when the temperatures of the junctions are t_1 and t_o. Similarly, the force in the part BBt_1At_2B is due to A and B, with temperatures t_1 and t_2 at the junctions; and that in the part BBt_2At_3 is due to A

and B with temperature t_2 and t_3 at the junctions. Hence the electromotive force, due to temperatures t_3 and t_o, is equal to the algebraic sum of the electromotive forces due to temperatures t_3 and t_o, t_2 and t_1, t_3 and t_2, respectively.

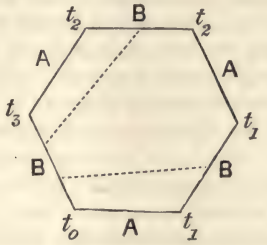


FIG. 179.

This result is quite general, and, therefore, *the algebraic sum of the various electromotive forces in a compound circuit, which is composed of a number of pieces of two metals with their junctions at various temperatures, t_o and t_1, t_1 and t_2, ... t_{n-1} and t_n, is equal to the electromotive force in a two-junctioned circuit of the same metals with its junctions at the extreme temperatures t_o and t_n.*

We can thus obtain a comparatively large electromotive force by means of a small difference of temperature. This is the essential principle of the *Thermopile*, an instrument which, in conjunction with a galvanometer, is used for the measurement of small differences of temperature.

Lastly, arrange four metal wires, A, B, C, and B in the manner

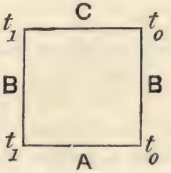


FIG. 180.

indicated in Fig. 180, and let the one wire B be raised to temperature t_1, while the other is kept at temperature t_o. The wires B merely serve as junctions, and so the total electromotive force is due to a circuit composed of A and C with the junctions at temperatures t_1 and t_o respectively. But we may join the two wires B by third wire of the same metal without altering the distribution of electromotive force in the circuit. And in the circuit Bt_1At_oBB the

electromotive force is due to A and B with junctions at temperatures t_1 and t_0; while in the circuit Bt_0Ct_1BB the electromotive force is due to B and C, with junctions at the same temperatures.

If we define the *thermo-electric power* of a circuit of two metals as the rate at which the electromotive force in that circuit varies per unit of difference in temperature between the junctions, we see that the result which we have just obtained shows that *at every temperature the thermo-electric power of* A *and* C *is the algebraic sum of the thermo-electric powers of* A *and* B, *and* B *and* C.

329. *Variation of the Electromotive Force with Temperature.*— Soon after Seebeck's discovery Cumming observed that in certain circuits (such as that of iron and copper), while one junction is maintained at a constant ordinary temperature and the other is gradually raised in temperature, the electromotive force gradually increases to a maximum, then diminishes, vanishes, and finally is reversed.

The law of variation of the electromotive force has been very fully investigated by Gaugain and others. They found that, with most pairs of metals, the curve which is obtained by plotting difference of temperature along the (horizontal) axis of x, and electromotive

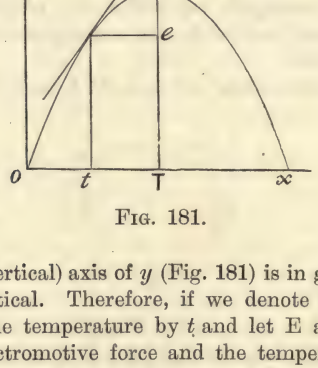


FIG. 181.

force along the (vertical) axis of y (Fig. 181) is in general a parabola with its axis vertical. Therefore, if we denote the electromotive force by e, and the temperature by t, and let E and T respectively represent the electromotive force and the temperature which correspond to the vertex of the parabola, we obtain

$$E - e = b(T - t)^2 \quad \ldots\ldots\ldots \quad (1)$$

where b is a constant; for this equation merely expresses the well-known property of the parabola, that the square of the distance of

any point on it from the axis is proportional to the distance of that point from the tangent at the vertex.

In particular cases the curve is a straight line; in others, it is made up of portions of parabolas with parallel (vertical) axes, but with their vertices alternately turned in opposite directions.

330. *The Thermo-electric Diagram.*—Now another well-known property of the parabola is that the rate of increase of the ordinate at any point per unit of increase of the abscissa is proportional to the value of the abscissa measured from the axis. The proof of this is simple, for (1) gives

$$\frac{de}{dt} = 2b(\mathrm{T} - t), \ldots\ldots\ldots\ldots (2),$$

which is the direct expression of the above statement.

If, therefore, we plot the value of de/dt (which is the thermo-electric power) against difference of temperature, we shall obtain, instead of a parabola, a straight line. And we may form a *self-consistent* diagram of such lines for any number of metals by means of observations on circuits consisting of each of these, in turn, with some standard metal whose line is made to coincide with the axis of temperature. (In the *true* diagram the lines might, of course, be

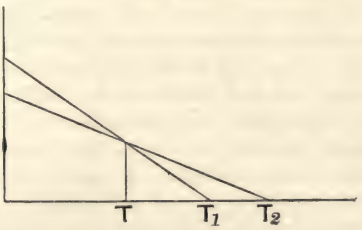


FIG. 182.

curves obtained from these straight lines by a process of shearing.) We see, by (2), that these various lines will intersect the axis of temperature at points which correspond to the temperatures of the maximum ordinates in the original diagram (Fig. 181). And, further, the point of intersection of any pair of lines in the diagram, indicates the temperature, T, at which the electromotive force in a circuit of the two corresponding metals attains a maximum value. This diagram is called the *thermo-electric diagram.*

Sir W. Thomson first suggested the construction of such a diagram. The actual construction of it, upon the assumption (suggested by

theoretical considerations) that the curves which represent the thermo-electric powers of the metals are, in general, straight lines, is due to Tait. The diagram on page 422 is reduced from his results.

The area included between two temperature lines and the lines which represent the thermo-electric position of any pair of metals represents the electromotive force in a circuit of these metals when the junctions are kept at the two given temperatures. This follows (see § 34) from the way in which the diagram of lines has been deduced from the diagram of parabolas.

331. *The Peltier Effect.*—A thermo-electric circuit forms a system which is in stable equilibrium. For, if it were in unstable equilibrium, increase of temperature of one of the junctions would produce effects which would still further increase the temperature. But we know that the application of heat to one junction produces a current of electricity which flows in a certain direction across that junction. Therefore by the principle of stable equilibrium (§ 15), we can assert that the passage of electricity in the given direction will cool the junction.

The current at the hot junction always flows from the metal which has the lower thermo-electric power to the metal which has the higher thermo-electric power. Conversely, heat is absorbed at a junction where a current flows in this direction, or is evolved at a junction where the current flows in the reverse direction. Peltier discovered this by direct experiments made without reference to any theoretical considerations; and so the phenomenon of the absorption or disengagement of heat at a junction across which electricity flows is known as the *Peltier Effect* at that junction. The total Peltier effect in any circuit vanishes when the two junctions are at the same temperature, for the absorption of heat at one of the junctions is equal to the disengagement of heat at the other.

332. *The Thomson Effect.*—In order to explain the fact that, in such circuits as iron-copper, the direction of the electromotive force changes when the hot junction is sufficiently raised in temperature, Thomson assumed that the Peltier effect vanishes at that temperature at which the electromotive force reaches its maximum value, that is, at the temperature at which the lines of the metals intersect in the thermo-electric diagram. The metals are then said to be neutral to each other, and so this temperature is called the *Neutral Temperature.*

Now no heat is being absorbed or developed at the junction which is at the neutral temperature, and heat is being developed at the cold junction, for there the current is flowing from the metal of higher thermo-electric power to the metal of lower thermo-electric

power. It would seem, therefore, that there is no source of thermal energy in the circuit by the transformation of which we can account for the development of electrical energy. But there is no other possible source of the electrical energy, and hence Thomson was led to predict that heat is absorbed at parts of the circuit other than the junctions, either in that metal in which the current flows from hot parts to cold parts, or in that metal in which it flows from cold parts to hot parts, or in both metals. And he subsequently verified his prediction by direct experiment.

In copper, heat is absorbed when the current is passing from cold parts to hot parts; in iron, it is absorbed when the current is passing from hot parts to cold parts. [It is assumed here that we know the direction in which a current is flowing. The convention by which this is determined will be stated in next chapter.] Now, in a tube through which a liquid is flowing, heat is absorbed by the liquid when it passes from cold parts to hot parts. Hence, in copper and similar metals, electricity acts as an ordinary fluid would do; and so Thomson speaks of the *specific heat* of electricity. It is positive in copper and similar metals, and is negative (at ordinary temperatures, at least) in iron and similar metals.

333. *Further Discussion of the Thermo-electric Diagram.*—The Peltier and the Thomson effects can be readily represented on the thermo-electric diagram.

Let e_1 represent the electromotive force in a certain circuit composed of some metal together with the standard one. T, being the neutral temperature, we get, at temperatures t and t' respectively, while the temperature t_0 of the cold junction remains constant

$$\mathrm{E}_1 - e_1 = b_1(\mathrm{T}_1 - t)^2$$

$$\mathrm{E}_1 - e'_1 = b_1(\mathrm{T}_1 - t')^2$$

whence

$$e'_1 - e_1 = 2b_1(t' - t)\left(\mathrm{T}_1 - \frac{t'+t}{2}\right) \ldots . (3).$$

Similarly, if we use any other metal giving the electromotive force e_2 with the standard metal under the same conditions

$$e'_2 - e_2 = 2b_2(t' - t)\left(\mathrm{T}_2 - \frac{t'+t}{2}\right) \ldots (4).$$

Hence the electromotive force in a circuit of the two given metals under the same conditions as to temperature is

$$\bar{e} = (e'_1 - e_1) - (e'_2 - e_2)$$

$$= 2(t' - t)\left[b_1\mathrm{T}_1 - b_2\mathrm{T}_2 - (b_1 - b_2)\frac{t'+t}{2}\right].$$

Now (2) shows us that $2b_1$ and $2b_2$ are the tangents of the angles at which the lines of the metals meet the line of the standard metal in

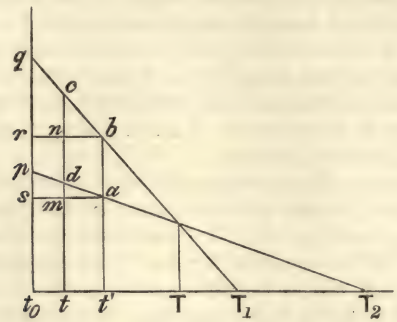


FIG. 183.

the thermo-electric diagram. So, if we let t_0 represent the absolute zero of temperature, while all the other temperatures are given in absolute units, we get

$$qt_0 = 2b_1T_1 \text{ and } pt_0 = 2b_2T_2$$

and so (Fig. 183) $qp = 2(b_1T_1 - b_2T_2)$. But $qp = 2T(b_1 - b_2)$,

whence $$b_1T_1 - b_2T_2 = T(b_1 - b_2),$$

and therefore $$\bar{e} = 2(b_1 - b_2)\ (t' - t)\left(T - \frac{t' + t}{2}\right), \ldots . (5),$$

which is the general expression for the electromotive force in a circuit of two metals in terms of the inclinations of their lines to the line of the standard metal, of their neutral temperature, and of the temperatures of the junctions.

Now (5) may be put in the form

$$\bar{e} = 2(b_1 - b_2)(T - t')(t' - t) + (b_1 - b_2)(t' - t)^2 \ldots . (6)$$

The first term on the right-hand side of (6) vanishes when $t' = t$ and also when the temperature of the hot junction is at the neutral point. It therefore represents the part of the electromotive force which corresponds to the Peltier effect. Therefore, if there is no other effect than the Peltier effect and the Thomson effect, the second term must represent the part of the electromotive force which corresponds to the Thomson effect.

If we suppose that unit quantity of electricity is transferred along the circuit under the electromotive force $\bar{e}$, the quantity $\bar{e}$ represents

the electric energy which is expended in the process, and therefore the quantity on the right-hand side of the equation expresses, on the one hand, the heat which is absorbed in the production of the electric energy, or, on the other hand, the heat which is evolved when unit quantity of electricity passes round the circuit under the electromotive force $\bar{e}$. The first term of (6) is therefore taken as the measure of the Peltier effect, while the second term is taken as the measure of the Thomson effect.

If the current flows round the circuit in the direction *abcda*, the area *abcda* represents the whole heat which is absorbed in the circuit during the passage of unit quantity of electricity.

But $2(b_1 - b_2)\ (T - t') = ab$, and therefore the area *abnma* represents the heat which is on the whole absorbed at the junctions. The whole area *abrsa* represents the heat which is *absorbed* at the hot junction, for at that junction the current is passing from the metal of lower thermo-electric power to the metal of higher thermo-electric power; while the area *msrnm* represents the heat which is *evolved* at the cold junction, for at the cold junction the current passes from the metal of high thermo-electric power to the metal of low thermo-electric power.

Again, $cn = 2b_1(t' - t)$, and therefore the triangular area $cnb = \frac{1}{2} \cdot 2b_1(t' - t)\ (t' - t)$ represents the heat which is *absorbed* in the metal of higher thermo-electric power. Similarly the area *amd* represents the heat which is *evolved* in the metal of lower thermo-electric power. [The former part has the positive sign prefixed to it in (6); the latter part has the negative sign prefixed.]

It is evident that we may state quite generally that heat is absorbed at any part of the circuit at which the current is passing from parts of lower thermo-electric power to parts of higher thermo-electric power, and is evolved at any part when the current is passing from parts of higher to parts of lower thermo-electric power.

The term $(b_1 - b_2)\ (t' - t)\ (t' - t)$ may be regarded as the product of the sum of the average specific heats of electricity in the various parts of the circuit into the range of temperature, each term in the sum being positive or negative according as it corresponds to absorption or evolution of heat. Therefore the form of the various terms $b_1 t'$, $b_1 t$, $b_2 t'$, and $b_2 t$, shows that the specific heat of electricity in any metal is proportional to the absolute temperature, and consequently the average which must be taken is the arithmetical mean; so that, if σ be the actual specific heat at temperature t, the average value throughout the range t will be $\frac{1}{2}\sigma$. Thus, when the electricity flows from b to c (Fig. 183), we may suppose that it flows along bq, and comes back along qc. In the first part of this process heat is

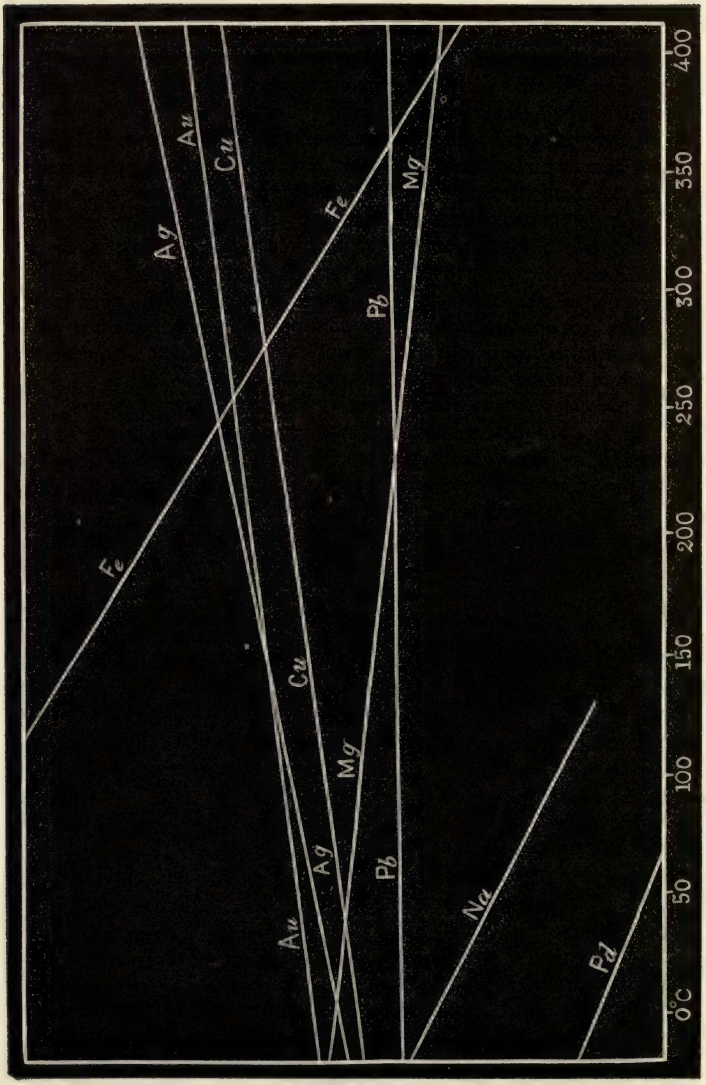

Ag
Au
Cu
Fe
Mg
Pb
Fe
Cu
Mg
Au
Ag
Pb
Na
Pd
0°C
50
100
150
200
250
300
350
400

FIG. 184.

absorbed; in the second part heat is evolved. In the first part the average specific heat is (say) $\frac{1}{2}k_1t'$, where k_1 is a constant. Similarly, in the second part, the average specific heat is $\frac{1}{2}k_1t$, which must be regarded as a negative quantity, since heat is being *evolved* as the electricity passes from cold parts to hot parts. The total sum for the two metals is therefore $\frac{1}{2}(k_1-k_2)\ (t'-t)=\sigma_1-\sigma_2$, so that the whole product $(b_1-b_2)\ (t'-t)\ (t'-t)$ is identical with $(\sigma_1-\sigma_2)\ (t'-t)$, and k_1 and k_2 are double of b_1 and b_2 respectively.

The specific heat of electricity in the metal whose line is $q\mathrm{T}_1$ (Fig. 183) is therefore represented by the line qr at the temperature t', and so on. It is negative or positive according as the line slopes downwards (like that of iron, Fig. 184) or upwards (like that of copper).

Le Roux found that the specific heat of electricity in lead is zero (or very nearly so), and therefore lead is chosen as the standard metal in the construction of the diagram.

Tait has found the very curious result that the specific heat of electricity in paramagnetic metals, such as iron and nickel, changes sign at least twice as the temperature is raised (see § 356).

CHAPTER XXIX.

ELECTRIC CURRENTS.

334. *Convection Current between Charged Conductors.*—A pith-ball placed between two oppositely charged insulated conductors, and free to move between them, will gradually destroy their electrification. When it comes in contact with the positively charged body, it receives a positive charge with which it moves, under the action of the electric force, towards the negatively charged body. It gives its positive charge to this body, and receives from it a negative charge, with which it again moves towards the positive conductor, and so on alternately.

If the charges of the two conductors were originally equal, the process will result in the complete destruction of their electrification. If only one of the two is originally charged, the process will result in the division of the charge between the two in the ratio of their capacities.

Positive electricity is carried in one direction, and negative electricity is carried in the reverse direction, by a convective process. The greater the electric force between the conductors is, the more rapid will be the motion of the ball, and the more nearly will the process approach to a continuous flow of electricity.

335. *Flow of Electricity in Metallic Conductors.*—If we place the two oppositely charged conductors in contact with each other by means of a metallic wire, a current of electricity will be established between them until their potentials are equalised. If one of the bodies be electrified, say positively, while the other is unelectrified, the result is (as above) that the charge is divided between the two bodies in the ratio of their capacities; and it is usual to say that positive electricity has flowed from the first body to the second, though the same result might have been produced by a flow of induced negative electricity in the opposite direction.

Any difference of potential between two parts of a conductor constitutes an electromotive force under which transference of

positive electricity will take place from the part at higher potential to the part at lower potential. And so long as the potentials are maintained constant, the quantity of electricity which flows from the one part to the other in a fixed time remains constant. The amount which flows per unit of time along the conductor from the one part to the other, is called the *Strength* or *Intensity* of the current.

The slightest difference of potential between two parts of a conductor will produce a current of electricity between these parts, but no current can be maintained without the expenditure of energy. That is to say, the flow of the current is opposed by a *Resistance* in the conductor. This is analogous to the flow of a liquid along a tube. A current is produced in the tube by the action of the slightest force, but work must be suspended in order to maintain the flow; for the motion is opposed by a resistance which is due to internal friction.

When an incompressible fluid flows through a tube because of a constant difference of pressure at its extremities, equal quantities of fluid pass every section of the tube in the same time. So in the flow of electricity along a conductor because of a constant difference of potential at its extremities equal quantities of electricity pass every section of the conductor in the same time.

When by any means a difference of potential is maintained at different parts of a conductor, we may draw in the conductor a series of equipotential surfaces. The electric stream lines are everywhere perpendicular to these surfaces.

Fig. 185 (*a*) represents the distribution of stream lines and equipotential lines in a thin circular conducting sheet, the centre of which is kept at a constant negative potential, while a point on its circumference is kept at an equal positive potential. The equipotential lines are in part open curves with their extremities on the circumference of the circular sheet, and in part they are closed curves surrounding the centre. If the whole diagram be revolved about the axis of symmetry, the various lines will trace out surfaces of flow and of constant potential in a conducting sphere, when the given points are maintained at constant (different) potentials.

In the upper half of Fig. 185 (*b*) the method by which the stream lines in the above case are drawn is exhibited. The circumference of the circle is divided into a whole number of equal parts and lines are drawn from A and C to the extremities of these parts. These lines are numbered from A round the circumference to the opposite end of the diameter AC.

The lower half of the figure exhibits the method of drawing the

equipotential lines when A and C are points in an infinitely extended conducting sheet. The nature of the difference between them and the equipotential lines of Fig. 185 (*a*) is apparent.

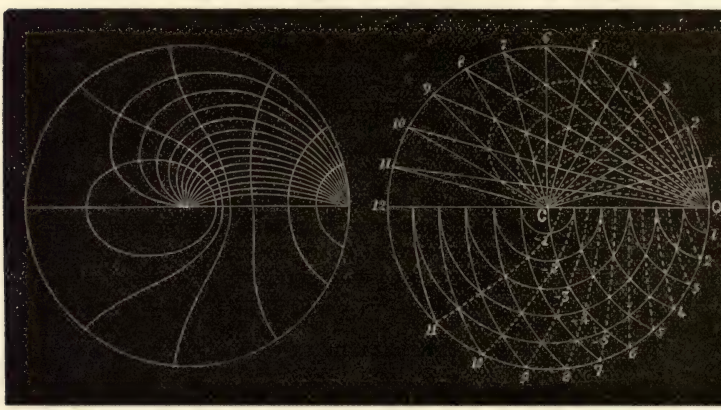


FIG. 185.

336. *Ohm's Law. Kirchoff's Laws.*—Ohm determined experimentally the relation which connects electromotive force, current-strength, and resistance in a conducting circuit. This relation is therefore known as Ohm's Law.

He found that in a conductor of given resistance the strength of the current is proportional to the electromotive force, and that in a conductor of variable resistance the strength of the current is inversely proportional to the magnitude of the resistance, if the electromotive force be constant. If E, C, and R represent respectively the electromotive force, the strength of the current, and the resistance, these results are expressed by the equation

$$E=CR$$

if we define unit current as the current which is maintained in a conductor of unit resistance by unit electromotive force.

This law enables us to calculate the resistance of a compound conductor composed of a number of conductors arranged in 'series,' that is, arranged so that the current flows from one to another of the several conductors in succession. Let E denote the total difference of potential in the circuit, and let R be the total resistance. Also, let e_1, e_2, etc., and r_1, r_2, etc., represent the corresponding

quantities for the several conductors in the circuit. The same current C flows through all the conductors, and hence the condition

$$E = e_1 + e_2 + \ldots .$$

gives, by Ohm's Law,

$$R = r_1 + r_2 + \ldots .$$

Therefore *the resistance of a number of conductors arranged in series is the sum of their separate resistances.*

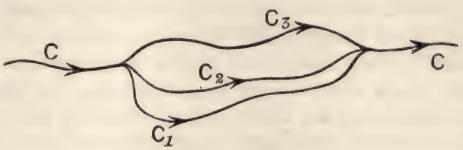


FIG. 186.

If the conductors be arranged in 'multiple arc,' as in Fig. 186, the condition

$$C = C_1 + C_2 + \ldots .$$

where C_1, C_2, etc., are the currents in the separate parts of the circuit, gives, by Ohm's Law,

$$\frac{E}{R} = \frac{E}{r_1} + \frac{E}{r_2} + \ldots ,$$

whence

$$\frac{1}{R} = \frac{1}{r_1} + \frac{1}{r_2} + \ldots .$$

The reciprocal of the resistance of a conductor is called its *conductivity*, and so this equation expresses the fact that *the conductivity of a compound conductor, formed of a number of separate conductors joined in the multiple arc arrangement, is the sum of the individual conductivities of these conductors.*

The law alluded to above, that the flow of electricity in conductors resembles that of an incompressible fluid, together with Ohm's Law, enables us to express the relations connecting electromotive, current-strength, and resistance in the various parts of any network of conductors, however complex. These laws, stated in the forms,

(1) The sum of the currents which flow towards any point of the network is equal to the sum of the currents which flow from it,

(2) The sum of the electromotive forces which act in any closed loop of the network is equal to the sum of the products of the current into the resistance in the several parts of the loop,

are known as Kirchoff's Laws.

337. *Electrolytic Conduction. Faraday's Laws.*—When electricity passes through certain conductors, chiefly liquids, decomposition of the conductor takes place; and it appears that the decomposition is a *necessary* accompaniment of the passage of the electricity. Such substances are called *electrolytes*, and the process of decomposition is termed *electrolysis*.

The current usually enters and leaves the electrolyte by metallic conductors, which are termed the *electrodes*, the conductor by which the current enters being distinguished as the *anode*, while the conductor by which it leaves is called the *cathode*.

The products of decomposition appear at the electrodes, the metallic constituent appearing at the cathode, while the other constituent appears at the anode. Thus, in the electrolytic decomposition of hydrochloric acid, hydrogen appears at the cathode, while chlorine is evolved at the anode. No decomposition occurs in the interior of the electrolyte. Hence it follows that when an electromotive force acts upon the electrolyte, the metallic constituent travels in the direction in which the current goes (that is, from high to low potential), and the other constituent travels in the opposite direction. The two constituents are therefore termed *ions*, the one which moves towards the cathode being called the *cation*, while the one which moves towards the anode is called the *anion*.

The laws of electrolytic conduction were fully investigated by Faraday. He found that the amount of the electrolyte which is decomposed by the passage of a certain quantity of electricity is strictly proportional to that quantity. The amount which is decomposed during the passage of a unit of electricity is called the *Electrochemical Equivalent* of the substance.

Faraday also found that the amount of any substance, which appears as anion or cation, is totally independent of the substance with which it is combined; and this shows that *the electrochemical equivalent of a substance is absolutely constant.*

The process of electrolysis has received extensive practical applications in the art of electrometallurgy.

338. *Polarisation. Ohm's Law in Electrolytes.*—The passage of electricity through a liquid results in the chemical decomposition of that liquid, and the electrical energy is transformed in part into potential energy of chemical separation.

Now, if two parts of a conductor are kept at constant (different) potentials, the work which is done in transferring unit quantity of electricity from the part at the higher potential to the part at the lower potential is equal to the difference of potential between the two parts—that is, to the electromotive force (E, say). Let H be the amount of heat which is developed in the combination of unit amount of the ions, from the state in which they are liberated, so as to form the compound electrolyte. This amount of heat is equivalent to the amount of work JH, where J is the dynamical equivalent of heat. Hence, in dynamical units, JqH is the work which is expended when unit amount of electricity passes through an electrolyte of which q is the electrochemical equivalent, and, therefore,

$$E = Jq\text{H}.$$

The quantities, J, q, and H, in this equation are all finite, and so E is finite; and, therefore, a finite electromotive force is required, in order to effect electrolytic decomposition.

If we keep the electrodes at an insufficient difference of potential, no decomposition can ensue; but if there be any difference of potential in the circuit, flow of electricity must occur. And we are here met with a difficulty, for Faraday's Law asserts that decomposition takes place in strict proportion to the amount of electricity which passes. But this difficulty can readily be explained. The phenomenon is precisely analogous to the charging of a Leyden jar. When the jar is charged under given conditions regarding potential, positive electricity flows into the one coating and negative electricity flows into the other until the potential of the jar is equal to that of the machine which is used to charge it. But the electrical energy is stored up in the dielectric which insulates the coatings of the jar, and will produce an equal reverse current of electricity whenever the coatings are put in metallic communication with each other, and we say that this reverse current is produced by the reverse electromotive force of the jar. Similarly, when an electromotive force, too feeble to cause decomposition, acts upon an electrolyte, flow of electricity takes place in the conducting parts of the circuit —that is, in the metallic and the electrolytic conductors. But no transference of electricity can take place between the electrode and the electrolyte, for decomposition must occur (by Faraday's law) if it did, and the electromotive force is not large enough to produce decomposition. The layer (of thickness comparable with molecular dimensions) between the molecules of the electrolyte and the molecules of the electrode acts as an insulator, and its two surfaces become oppositely charged, as those of a Leyden jar would, until

the reverse electromotive force produced in this way becomes equal to the direct electromotive force. If now the direct force be removed and the electrodes be joined so as to form a closed circuit, the reverse force will produce a reverse current.

This phenomenon is called *polarisation*, the reverse force is called the electromotive force of polarisation, and the reverse current is called the polarisation current.

Let E_o be the force which is required to produce decomposition. If E be the direct force, the difference $E - E_o$ is alone effective in producing a permanent current.

The quantity E_o is not constant except under fixed conditions; and, in practice, the conditions under which the decomposition takes place vary greatly. The electrodes or the electrolyte may have some chemical or molecular attraction for the products of decomposition, and the electromotive force which is required to produce the decomposition will be less in proportion as this attraction is greater. As an extreme case, an infinitely small electromotive force would effect the decomposition of an electrolyte, the constituents of which had no greater attraction for each other than they had for the substance of the electrode. Again, if the products of electrolysis are gaseous, and are dissolved by the electrolyte, decomposition will take place more readily than it would if the gases were not dissolved. When the process is first started the gases may dissolve readily, but their solution will occur with greater and greater difficulty until saturation is reached, when it attains a maximum. Further, if the gases evaporate from the electrolyte, saturation will never be fully attained, and the maximum attainable state of saturation will depend upon the partial pressure of the gases in the atmosphere which is in contact with the electrolyte.

These and other causes which affect the electromotive force have been fully discussed by Von Helmholtz, who has given a complete thermodynamical theory of polarisation, the results of which accord very well with experimental facts.

If n be the number of atoms in the electrochemical equivalent of a substance, the quantity q/n may be regarded as the 'atomic charge,' which is constant since q and n are constant. This constancy of the atomic charge points to an intimate relation between electricity and matter, and the fact that the electrochemical equivalent of a substance is constant whether it be electrolysed from combination with a monad, dyad, or triad, etc., element shows that the atomic charge of a dyad element is double that of a monad element, that the atomic charge of a triad element is three times that of a monad element, and so on. This led Von Helmholtz to

regard chemical affinity as the result of a greater attraction of some atoms than of others for positive or negative electricity. Thus, for example, oxygen has a greater attraction for negative electricity than hydrogen has, while hydrogen attracts positive electricity more strongly than does oxygen. And Von Helmholtz regards the affinity of oxygen and hydrogen for each other as the result of the electrical attraction between positively charged hydrogen and negatively charged oxygen.

The atomic charge is excessively small since n is very large; but, since the atoms are at an exceedingly small distance apart, the attraction may be very great.

These considerations enabled Helmholtz to explain a phenomenon which appeared to be at variance with Faraday's law. An extremely feeble constant current, inappreciable to all but the most delicate instruments, flows in an electrolytic circuit under electromotive forces far too feeble to produce decomposition. Von Helmholtz showed that this was due to the presence of dissolved gases. Thus oxygen in solution will gradually find its way to the negative electrode and will receive a negative charge. It then becomes subject to the electromotive forces and travels to the positive electrode, to which it gives up its negative charge. In this way a constant 'convective' current of electricity is kept up, and Faraday's law is not subject to exception.

Helmholtz observed the existence of this convective current even when he had carefully freed the electrolyte from dissolved gases by the most perfect processes. Here again there is no real exception to Faraday's law, for Helmholtz's thermodynamical theory shows that some dissolved gases must be present if the (liquid) electrolyte is to be in stable equilibrium; and, therefore, even if all dissolved gases were abstracted, decomposition of the liquid would go on until the requisite minimum of dissolved gases products were present.

In the case of electrolytes, Ohm's law must be put into the form $E - E_0 = CR$, E_0 being, as above, the electromotive force of polarisation. Its applicability has been verified, in all cases, within the limits of experimental error. The variability of E_0 renders the proof of the law extremely difficult.

339. *Production of Electric Currents.*—An electrolytic circuit constitutes a system which is in stable equilibrium, for a finite electromotive force is required in order that decomposition may take place, and the removal of the electromotive force causes the cessation of the decomposition. Consequently, we conclude (§ 15) that, in this stable system in which the passage of a direct current produces

chemical separation, the energy of chemical separation will be transformed into the energy of a reverse electric current, when the conditions are such that this reverse current can flow.

Therefore, in a closed conducting circuit, which includes an electrolyte, and which has associated with it energy of chemical separation, the sum of the electromotive forces may not be zero, but will produce an electric current, while chemical combination proceeds. The sum of the forces can differ from zero only because there is a chemical source for the energy of the current which the resultant force produces; just as, in a thermo-electric circuit, the electric energy is derived from the heat energy which is supplied.

An arrangement of this kind may therefore be used for the production of an electric current; and it is of no consequence whether the chemical separation is produced by the previous passage of an electric current through the circuit under the action of an external electromotive force, or is given as an initial condition produced by chemical processes or otherwise.

An arrangement of the latter class constitutes a *Primary Cell;* one of the former class constitutes a *Secondary Cell.*

The equation in § 338 indicates a method of determining the electromotive force in any such circuit; for it asserts that 'the electromotive force of an electrochemical apparatus is in absolute measure equal to the mechanical equivalent of the chemical action on one electrochemical equivalent of the substance.' This was first pointed out by Sir W. Thomson. Conversely, the energy which is developed in a given chemical action can be estimated by means of measurements of the electromotive force of an electrochemical apparatus in which that action is developed.

340. *Primary Cells.*—Primary cells are sub-divided into two classes according as one fluid or two fluids are used in their construction.

All the older cells were of the one-fluid type, and the conducting circuit was generally made up of a plate of zinc, a plate of copper, and sulphuric acid. The sulphuric acid, combining with the zinc, produces zinc sulphate, and hydrogen is liberated at the copper electrode. Positive electricity flows from the zinc to the copper through the liquid (Fig. 187).

Such a cell has many disadvantages. Slight impurities in the zinc (possibly even differences in its physical constitution) at different parts give rise to local currents of electricity between these parts, and thus lead to solution of the zinc without the production of effective currents in the main circuit. This local action may be

largely prevented by amalgamating the surface of the zinc, for the surface is thus rendered more nearly homogeneous. Again, the evolution of hydrogen at the copper electrode gives rise to a reverse

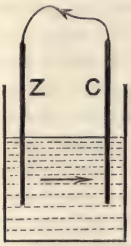


FIG. 187.

electromotive force of polarisation. And, ultimately, when the sulphuric acid is transformed into zinc sulphate, zinc, instead of hydrogen, is deposited upon the copper plate. When this process is complete, both electrodes are practically composed of zinc; and, from the symmetry of the arrangement, it is evident that the sum of the electromotive forces in the circuit must be zero, so that the current ceases.

Daniell improved this cell by the introduction of two fluids, separated by a porous diaphragm which did not prevent the passage of electricity. A zinc rod Z (Fig. 188) is placed inside a porous cell,

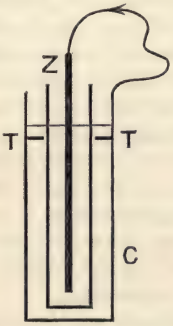


FIG. 188.

which is contained in an outer copper vessel C. The porous cell contains a solution of zinc sulphate, and the copper vessel contains a saturated solution of copper sulphate. In order to keep this

solution saturated, crystals of the sulphate of copper may be placed on a perforated tray T, inside the copper vessel. These crystals are dissolved away when the solution becomes weakened.

When the cell is in action, the current passes from zinc to copper through the liquid. The zinc sulphate is electrolysed into zinc and 'sulphion' (SO_4). The SO_4 acts upon the zinc and forms zinc sulphate. Thus the zinc sulphate always tends to be saturated. The copper sulphate is simultaneously electrolysed into copper and sulphion, and the copper (being the metallic ion, and therefore travelling with the current, which is assumed to travel in the direction in which positive electricity goes) is deposited upon the copper vessel. The zinc, which results from the electrolysis of the zinc sulphate, is not deposited at all, but unites with the sulphion produced by the electrolysis of the sulphate of copper, and so reforms zinc sulphate, which remains inside the porous cell. Thus hydrogen is not evolved at the copper electrode, and polarisation is reduced to a minimum.

If the current be too strong, or if the copper sulphate be not saturated, some hydrogen will be evolved.

Daniell's cell gives an extremely constant electromotive force if the conditions which are necessary to its proper action are maintained.

Bunsen's cell is also a two-fluid cell. A rod of carbon is placed in strong nitric acid, which is contained in the porous cell, and a zinc plate is placed in an outer cell of glazed earthenware which contains an aqueous solution of sulphuric acid. Hydrogen is evolved at the carbon, but is at once oxidised by the nitric acid. The fumes which are given off from the acid are very poisonous, so that these cells should be kept in a separate room, as far as possible, if a considerable number of them are in use.

Grove's cell resembles that of Bunsen in its chief features. The carbon is replaced by a sheet of platinised silver. The process of platinisation gives a large, very finely corrugated surface, which favours the diminution of polarisation.

The electromotive forces of the two latter cells are very nearly equal to each other, and are considerably larger than that of a Daniell cell, but are not so constant.

The single-fluid *bichromate cell* (bichromate of potassium dissolved in an aqueous solution of sulphuric acid, into which dip zinc and carbon plates) gives a high electromotive force, and is greatly used when a powerful current is required for short periods.

The *Léclanché cell* consists of a porous cell filled with manganese dioxide, in which a carbon rod is placed. A zinc rod is placed in a

solution of ammonium chloride, outside the porous cell. The electromotive force of this cell rapidly diminishes when it is made to produce a current; but it soon attains its original value after the current is stopped. Its great advantage is that it remains in good working order for months.

Many other forms of cells, including modifications of the above, are in practical use; but the present examples will serve sufficiently as illustrations of their general principles.

The zinc and the carbon, or the zinc and the copper, etc., are called the *elements* of the cell; carbon and copper being positive elements, while zinc is a negative element (for the current flows from carbon or copper to zinc through the external conductor).

Various cells may be joined together to form a voltaic *battery*. When the positive element of one cell is joined to the negative element of the next, the cells are said to be coupled in *series;* or to be coupled for electromotive force, for, in this case, the electromotive force of the battery is the sum of the electromotive forces of the several cells. When all the positive elements are joined together, and all the negative elements are joined together, the cells are said to be coupled for *quantity*, for the whole battery acts as one large cell of the same kind, and produces a powerful current.

341. *Secondary Cells.—Grove's Gas Battery* was the earliest form of secondary cell. It consists essentially of two glass tubes (A, B, Fig. 189), which are fitted into the two necks of an ordinary Woulfe's bottle. These tubes are open at their lower ends, but are

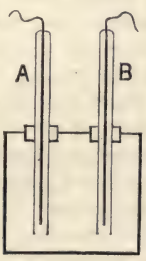


FIG. 189.

closed at their upper ends. Platinum wires are inserted through the glass at the closed ends, and are welded to strips of platinum foil which extend throughout nearly the whole length of the tubes. The tubes and the bottle are filled with (say) dilute solution of sulphuric acid.

If a current be sent round the circuit from A to B through the liquid, oxygen will be evolved in A and hydrogen will be evolved in B; and a reverse electromotive force will be produced. If the direct current be stopped, and if the wires let into A and B be joined, this reverse force will produce a reverse current setting from B to A through the liquid. The liquid will be decomposed, oxygen tending to appear in B and hydrogen tending to appear in A. The gases are not really set free; they combine with the gases already in the tubes to form water; and this process goes on until both tubes are again filled with the liquid solution.

[It is interesting to note that the arrangement may be used as a primary battery. Instead of evolving the gases by a direct current, we may introduce them independently into the tubes, and the current will proceed from B to A through the liquid as formerly.

The current will flow even if hydrogen be introduced into B while A is left full of liquid. The hydrogen in B unites with the oxygen which is produced by the electrolytic action of the current, and hydrogen is evolved in A. The sum of the electromotive forces in the circuit soon becomes zero, since there is hydrogen in both tubes, and the current stops.

If A be filled with ordinary oxygen while B is left full of liquid, very little effect will be observed; and this shows that electrolytically evolved oxygen differs considerably in its properties from ordinary oxygen. If A contains ozone, the action will proceed as before.]

Planté's secondary cell consists of two large sheets of lead, separated by a sheet of guttapercha, and rolled up into a spiral form. These sheets are placed in a vessel which contains dilute sulphuric acid, and an electric current is passed through the cell from one sheet of lead to the other. The oxygen which is evolved at the one sheet forms a layer of peroxide of lead on its surface. The direct current being stopped, and the terminals of the lead sheets being joined, a reverse current flows through the cell. The hydrogen which is now evolved partly reduces the peroxide of lead, and the oxygen unites with the lead of the other electrode to form peroxide; and, when both electrodes become similar, the current stops. A current is now passed, from an external source, through the cell in the same direction as that in which the reverse current flowed. When the gases bubble off freely at the electrodes, this current is stopped, and the polarisation current is allowed to flow; and so on alternately for a number of times. This process, which is said to *form* the cell, gradually changes the lead into a spongy condition, and thus greatly increases the charge which the cell can contain.

When the process of formation is complete, the cell is always charged by a current in one direction only.

Faure's cell is essentially similar to Planté's, but the plates of lead are at first covered with a coating of a lower oxide of lead. When the direct current is passed, the oxide on the one plate is changed to the peroxide, while that on the other is reduced to the metallic condition. This avoids the tedious process of forming the cell by alternate currents in opposite directions.

The modern *accumulator* is constructed on essentially the same principles as the Faure cell is constructed on, though various improvements are introduced. Its electromotive force exceeds that of a Bunsen cell, while its resistance is extremely small, so that it is capable of producing a very powerful current.

342. *Transformations of Electric Energy in Conducting Circuits.*—If the potential V of a conductor be kept constant while its charge alters by the amount Q, the electric energy changes by the amount VQ. Hence, if the quantity Q of electricity flows from a part of a conductor where the potential is V to a part where the potential is V′, the energy expended in the production of this flow is $(V - V')Q$, that is, it is EQ, where E is the electromotive force which acts between the given parts. Hence, if a current of intensity C is maintained between these parts, the rate at which energy is expended in the process is EC.

This energy, which is associated with the current, will be transformed into heat in the circuit, if no other transformations take place. If a reverse electromotive force E_0 acts in the circuit, work is expended at the rate E_0C in making the current flow in a direction opposite to that in which the reverse force acts. And this energy is transformed into heat at the place where the reverse force acts. (The Peltier evolution of heat at a thermo-electric junction is a case in point.)

If no part of the direct electromotive force acts against a reverse force, and if R is the resistance of the circuit, Ohm's law gives $E = CR$, whence we see that the rate at which heat is developed in the circuit is, in electrical units,

$$RC^2.$$

that is to say, *the rate at which heat is developed in a circuit is directly proportional to the resistance of the circuit and to the square of the intensity of the current which flows through it.* This is known as Joule's Law.

Part of this heat is produced in the cell which is used for the production of the current, but by making the external resistance very large in comparison with the internal resistance of the cell, this

portion may be made as small a fraction of the total amount as we please.

This principle is utilised in the process of electric lighting. The carbon filament of an 'incandescent' lamp is of relatively large resistance. Almost all the heat into which the electric energy is transformed is developed in the carbon, which becomes 'white-hot.' In the 'arc' lamp the chief part of the resistance is in the air-gap between the carbon poles, and the heat which is developed raises the air to so high a temperature that it becomes intensely luminous. [Air at ordinary temperatures is an insulator, but hot air admits of the passage of electricity through it with comparative ease. The heating of the air is effected by allowing the carbons to touch each other, so that the current flows through them and makes them red-hot near the point of contact. When this results, the carbons may be separated to a slight extent, and the current will continue to flow.]

The energy of an electric current may be directly transformed into mechanical work by means of an electro-motor (§ 366).

343. *Measurement of Electromotive Force, Current Strength and Resistance.*—The electromotive force which acts between two parts of a conductor may be determined directly by means of the electrometer (§ 325), for the problem is merely one of the determination of difference of potential.

The strength of the current which flows in a circuit may be found directly by placing an electrolytic cell (or voltameter) in the circuit. The quantity of electricity which passes through the cell per unit of time is, by Faraday's law, directly proportional to the amount of chemical decomposition which takes place per unit of time. [The arrangement which was described as Grove's gas battery, in § 341 constitutes a voltameter, and the amount of oxygen or of hydrogen which is evolved per unit of time in either inverted tube can be measured directly. If Q be this quantity, while q is the electrochemical equivalent of the substance, the strength of the current is Q/q.]

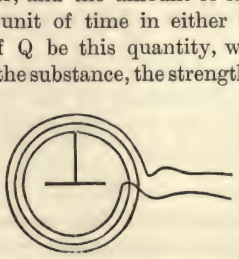


FIG. 190.

In practice the *galvanometer* is more frequently used for the measurement of the strength of a current. It consists essentially

(Fig. 190) of a coil of wire within which a magnet is freely suspended. The magnet lies normally in the plane of the coil, but when a current flows round the coil, the magnet tends to place its length at right angles to the plane of the coil. This tendency is opposed by the action of a constant external magnetic force, and the tangent of the angle of deflection of the magnet is proportional to the strength of the current (§ 369).

The resistance of a wire may be determined in terms of any given unit by means of the arrangement known as *Wheatstone's Bridge*. Let four conductors, the resistances of which are r_1, r_2, r_3, and r_4, be arranged as in Fig. 191, and let a battery b be placed between the points A and B, so that currents flow (say) in the way which is indicated by the arrows. The potential at the point C is inter-

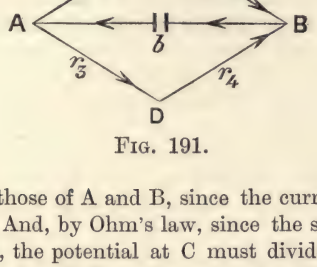


FIG. 191.

mediate between those of A and B, since the current flows, through C, from A to B. And, by Ohm's law, since the same current flows along AC and CB, the potential at C must divide the difference of potential between A and B in the ratio of r_1 to r_2. Similarly the potential at D will divide that difference in the ratio of r_3 to r_4. Hence, if C and D are at the same potential (which may be determined by the fact that no current will then flow through a galvanometer which is placed between C and D), the resistances must be connected by the relation

$$\frac{r_1}{r_2} = \frac{r_3}{r_4}.$$

And consequently, if we know the ratio r_3/r_4, and also the absolute value of r_2, we can calculate the value of r_1.

If we know the value of any two of the three quantities—electromotive force, current-strength, and resistance—we can calculate that of the third by means of Ohm's law (E=CR). Or, if H be the heat which is developed in a conductor by transformation of electric energy, while J is the mechanical equivalent of heat, we can calcu-

late the values of any two of the three quantities, when we know that of the third, by means of Joule's law,

$$EC = RC^2 = JH.$$

The resistance of a metallic conductor increases when its temperature is raised, but the resistance of an electrolytic conductor diminishes when its temperature increases. The law of variation being known, we can determine temperature by means of measurements of resistance. This method is very useful when the temperature is high.

A discussion of the various units in terms of which electrical quantities are measured will be given in Chap. XXXI.

The variation of resistance with temperature forms the basis of the most delicate method for the measurement of radiant energy. The *bolometer* which is used for this purpose consists essentially of an extremely sensitive and well-balanced Wheatstone's Bridge. This bridge is thrown off balance, and a current is produced through the galvanometer when radiant heat falls on one of the resistances.

CHAPTER XXX.

MAGNETISM.

344. *Fundamental Phenomena.*—Certain bodies, when they are suspended in such a way as to be able to turn in any direction, are found to have a marked tendency to place a definite set of lines drawn in their substance parallel to a definite direction in space. Such bodies are said to be *magnetised*, and are called *magnets*.

This property of magnetisation is possessed notably by one of the oxides of iron (lodestone), but it may be induced to a much greater extent in pieces of steel or metallic iron. The metals cobalt and nickel are also capable of becoming strongly magnetic. All other substances have relatively extremely feeble magnetic properties.

Magnets are classified as *permanent* or *temporary*, according as they do or do not retain, in large part at least, their state of magnetisation after the removal of the influence which caused it to be manifested. A bar of steel is of the former kind, a bar of soft iron is of the latter kind.

Any magnetic substance, when it is placed in the immediate neighbourhood of a magnetised body (more generally in a field of magnetic force, see § 362), becomes magnetised and retains its magnetisation so long as it remains in that position. Whether or not it will retain its magnetisation after removal from the neighbourhood of the magnetised body depends upon its physical constitution and upon circumstances which will be considered afterwards. And further, the substance in which magnetisation is thus induced will (in general) be attracted by the magnetised body.

For the sake of definiteness, let us consider the action of an ordinary 'bar' magnet, *i.e.*, a permanent magnet made of a rectangular or cylindrical bar of steel. If all similar parts of this bar are similarly magnetised (a condition which is not realisable in practice), or if it be symmetrically magnetised with regard to its axis of figure, it will, when freely suspended, place its axis of figure in the definite direction in space above alluded to. One definite end of the magnet will point, on the whole, northwards, the other will

necessarily point southwards. [The magnet, if turned round exactly end for end, may remain in the reverse position for a brief time; but it is essentially in unstable equilibrium, and will, if disturbed to the slightest extent from rest, turn round into its normal position.]

The suspended magnet will depart from this normal attitude if we bring up another magnet into its neighbourhood. Those ends of the two which naturally point northwards appear to repel each other. Those which point southwards also exhibit mutual repulsion; whilst those which naturally point oppositely appear to attract each other.

345. '*North*' *and* '*South*' *Magnetism.*—The phenomena which we have just considered present obvious analogies to electrostatic phenomena. Two electrostatic systems, each consisting of two oppositely charged, insulated, rigidly-connected, conducting spheres, placed near each other in a uniform field of electrostatic stress, would exhibit similar mutual action, and would, when free from each other's influence, take up a position in which the line joining the centres of the insulated spheres coincided in direction with the lines of electrostatic force in the surrounding medium. Hence, by analogy, we may assume that there are two kinds of magnetism; that like kinds repel each other; that unlike kinds attract each other; and that the force of attraction or repulsion diminishes as the distance between the attracting or repelling bodies increases.

It is usual to call the magnetism, which is found at that end of a magnet which points northwards, *north* magnetism; while the opposite kind is called *south* magnetism. But, since it is usual to distinguish the ends of a magnet by colouring the north-pointing end red and the south-pointing end blue, the terms *red* and *blue* magnetism are sometimes used instead of these, though their use cannot be commended.

Further, just as the action of an electrified body separates the neutral electricities in an adjacent conductor, we might expect by analogy that a magnetic substance would become magnetised so long as it remained in the neighbourhood of a magnet; and that it would be attracted towards the magnet just as the conductor would be attracted to the electrified body. All these results happen, but it is not well to push the analogy too far. Thus, while the conductor ceases to exhibit electrification when it is removed from the influence of the electrified body, a magnetic body will not necessarily (or even generally) cease to exhibit magnetisation when it is removed from the influence of the magnet. And it must be remembered that the phrase 'two kinds of magnetism' is merely adopted as a matter of convenience. (*Cf.* § 308.)

346. *Paramagnetic and Diamagnetic Bodies.*—Another point in which the direct analogy between electrical and magnetic action breaks down is in the *repulsion* of some bodies from a magnet. Let us suppose that the magnet is so long that the magnetic body which we are considering is subject only to the action of the magnetism at the near end of the magnet. Under this condition some bodies are attracted to the magnet, while others are repelled from it. Bodies of the former kind are called *paramagnetic* bodies; those of the latter kind are called *diamagnetic* bodies.

There is nothing in electricity corresponding to diamagnetism.

347. *Magnetism a Molecular Phenomenon.*—The great distinction between electrical and magnetic phenomena lies in the absence of anything of the nature of *conduction* of magnetism. While it might seem that the disappearance of induced magnetisation, when a piece of soft iron is withdrawn from the neighbourhood of a magnet, is due to the flowing together of the two opposite kinds of magnetism, the persistence of magnetisation to an appreciable extent when the soft iron is replaced by hard steel at once disposes of this view.

And, further, if we bring a magnetic substance into contact with one end of a magnet and then withdraw it from contact, no interchange of magnetism takes place, though an interchange of electricity would occur if the substances were electrified conductors. Also, while a conductor under the influence of an electrified body may be divided into two oppositely charged portions, it is impossible to divide a magnet into two oppositely magnetised portions—that is to say, it is impossible to isolate one kind of magnetism.

Every portion, however small it may be, into which a magnet may be broken exhibits properties precisely similar to those which were manifested by the complete magnet. We conclude, therefore, that this would still hold if the magnet were reduced to its constituent molecules; that each molecule of a magnetised body is itself a little magnet.

It is easy to explain, upon this assumption, how it is that magnetisation is only evident near the ends of a magnet. For, if the

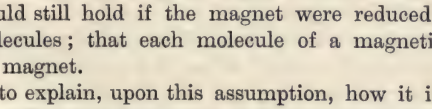


FIG. 192.

little circles in Fig. 192 represent the magnetised molecules, we see that the effect of any north end of a molecule at external points is

counterbalanced by the effect of the south end of an immediately adjacent molecule. It is only at the extremities of such a chain of molecules that the magnetisation can become manifest through the production of external effects, and the magnetism is of opposite kinds at the two extremities of the chain.

348. *The Law of Magnetic Attraction and Repulsion.*—We can investigate, by methods to be discussed subsequently (§ 358), the law of attraction or repulsion between the quantities of magnetism which we assume to exist at the ends of magnets. The results of such measurements make it evident that the force between two quantities is directly proportional to the magnitude of each quantity, and is inversely proportional to the square of the distance by which they are separated. If we choose to regard an attractive force as negative, and a repulsive force as positive, we can symbolise this law by the equation

$$F = \frac{qq'}{s^2},$$

where q, q', represent the quantities of magnetism, and s is the distance between them; for F is positive or negative according as q and q' are of like or of opposite signs.

This law is identical in form with the law of electrical attraction or repulsion, and hence all the results which we have previously deduced in the theory of electrostatics are capable of direct application in magnetostatics.

349. *Poles, Axis, and Magnetic Moment of a Magnet.*—Those two points of a magnet, at which its north and south magnetisms may be supposed to be concentrated, in order to produce the same effects at external points as the actual distribution of magnetism produces, are called the *Poles* of the magnet; and the line joining the poles is called its *Axis*.

In the case of a uniformly magnetised (rectangular or cylindrical) bar magnet, the poles would be at the geometrical centre of the ends of the magnet, and the axis would coincide with the axis of figure. In any actual magnet the poles are not exactly at the ends.

The quantity of north magnetism at the one pole of a magnet, or the (equal) quantity of south magnetism at the other pole, is called the *Strength* of the pole. The product of the strength into the distance between the poles is called the *Magnetic Moment* of the magnet. It is obviously analogous to the moment of a couple (§ 70).

350. *Lines of Magnetic Force. Magnetic Potential.*—A region

throughout which magnetic force is manifested is called a *Field of Magnetic Force*, or, more shortly, a magnetic field. And we may imagine this field to be filled with lines of magnetic force drawn in the direction in which a north pole would be moved. If we draw from any magnetic pole a number of lines of force numerically equal to 4π times the strength of the pole, the number of these which cross unit of area of any plane surface passing through any point can be made to represent the strength of the field—*i.e.*, the magnitude of the force—at that point in the direction of the normal to the given plane. In fact, as we have already seen, all the results previously given regarding electric lines of force can be at once applied to magnetic lines of force, and we, therefore, do not require to repeat them here. It is merely necessary to replace the term 'electrified body' by the term 'magnetised body,' the term 'positive electricity' by the term 'north magnetism,' 'negative charge' by 'quantity of south magnetism,' and so on.

A line of force can be readily traced out by means of a very small magnet, freely suspended, which is always moved in the direction in which it points. In every position its length is tan-

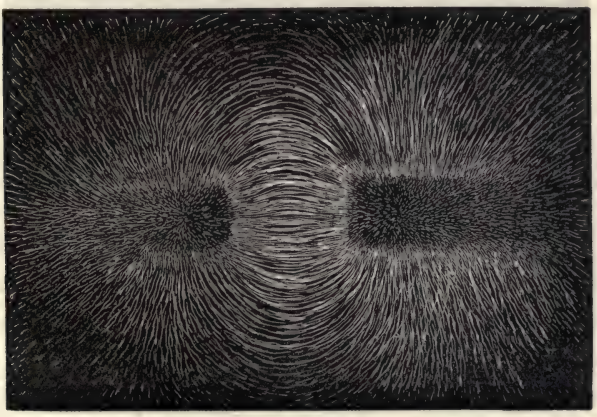


FIG. 193.

gential to the line of force which passes through its centre. The lines of force due to any group of magnets can also be readily shown by means of iron filings dusted over a sheet of paper, which is placed over the magnets. The filings become magnetised and turn so as to place their lengths in the direction of the force. A slight

tapping of the paper will cause the filings to group themselves in definite lines, each of which coincides with a line of force. The vibration of the paper throws the filings up into the air for a moment, so that they are free to accommodate themselves to the influence of neighbouring filings. (See Figs. 193, 194.)

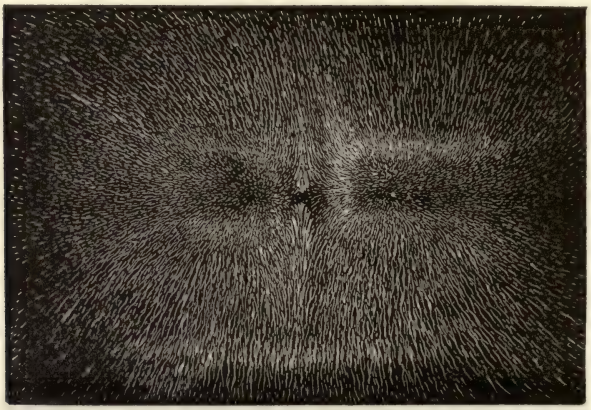


FIG. 194.

Following the electrostatic analogy, we may define the *Magnetic Potential* at any point, due to a magnetic pole, as the work which is expended in bringing a unit north pole—that is, a north pole of unit strength—from an infinite distance to that point. The results already deduced regarding electrostratic potential will then apply directly to our present subject.

351. *Magnetic Intensity. Magnetic Induction.*—If a rectangular bar-magnet were uniformly magnetised in the direction of its length (say), it is obvious that its total magnetic moment is equal to the sum of the moments of any number of parts (uniformly magnetised in the same manner), into which we may suppose it to be divided. For, if $L=l_1+l_2+\ldots+l_n$ be the total length of the magnet, and if Q be its pole-strength, we have

$$LQ=(l_1+l_2+\ldots+l_n)Q=l_1Q+l_2Q+\ldots+l_nQ,$$

which proves the proposition so far as transverse division is concerned. And, since the magnetisation is uniform, if we divide it longitudinally, each part becomes a magnet, whose pole-strength is

proportional to the area of its end. Hence, if $A = a_1 + a_2 + \ldots . a_n$ be the total area of the end, and if $\mathfrak{I}$ be the strength per unit of area, so that $\mathfrak{I}A = Q = \mathfrak{I}(a_1 + a_2 + \ldots . + a_n) = q_1 + q_2 + \ldots . + q_n$ where q_1, etc., are the strengths of the several parts, we have

$$LQ = Lq_1 + Lq_2 + \ldots . + Lq_n,$$

which proves the statement for longitudinal division.

But it is obvious that the statement is true whatever be the forms of the parts into which the magnet is divided, for each part may be supposed to be built up of an infinitely great number of infinitely small rectangular portions. And it follows from this consideration also that the proposition is true whatever be the form of the original magnet.

The quantity $\mathfrak{I}$, the pole-strength per unit of area, is called the *Intensity of Magnetisation* of the given magnet. It is evident that we may regard it as being the magnetic moment per unit of volume.

Let us imagine a cylindrical crevasse to be cut out in the interior of a uniformly magnetised body. Let it be bounded by plane surfaces perpendicular to the direction of magnetisation; and, while all its dimensions are infinitely small, let the perpendicular distance between these planes be infinitely smaller than the transverse dimensions. The surface density of magnetism on the plane faces of the cavity is $\mathfrak{I}$, north magnetism being distributed on the plane face next the south pole of the magnet, while south magnetism is distributed on the other; and hence the force in the space between the planes is $4\pi\mathfrak{I}$ (§ 99). We may therefore suppose that $4\pi\mathfrak{I}$ lines of force are drawn per unit of area across this cavity in the direction of magnetisation (that is, from the south pole to the north pole within the substance of the magnet). It is customary to call these lines the *Lines of Magnetisation.*

The lines of magnetisation do not constitute all the lines of force in the interior of the magnet. There may be lines of force due to external magnetisation. This distribution of force must be investigated in precisely the same manner as that in which we investigate the distribution of force outside a magnet.

Let us suppose that the cylindrical cavity is infinitely long in comparison with its cross dimensions. The magnetisation at the ends of this cavity exerts no effect upon a point at its centre, and hence any force found at the centre must be due to external magnetisation. This quantity is denoted by the symbol $\mathfrak{H}$, and is called the magnetic force at the point. The total force, $\mathfrak{B}$, is called the

Magnetic Induction at the given point in the magnet, and is equal to

$$4\pi\mathfrak{k}+\mathfrak{P}.$$

It is usual to call the total lines of force inside a magnet, *Lines of Induction.* They consist, therefore, partly of lines of magnetisation, and partly of the lines of force within the uncut magnet. They are continuous with the lines of force external to the magnet.

[It must be remembered that the three quantities $\mathfrak{B}$, $\mathfrak{k}$, and $\mathfrak{P}$, are vector quantities, and are therefore subject to the laws of vector addition (§ 40). In most practical cases, however, $\mathfrak{k}$ and $\mathfrak{P}$ are either similarly or oppositely directed.]

The force due to the magnetism at the surface of the magnetised body is included in the quantity $\mathfrak{P}$, and is obviously directed oppositely to $\mathfrak{k}$ since it acts in the direction of a line drawn from the north pole to the south pole through the material of the body. It therefore acts so as to demagnetise the body, and has its greatest value $2\pi\mathfrak{k}$ at points close to the ends of the magnet (Compare § 99). [To obviate demagnetisation, bar magnets, when not in use, are placed parallel to each other with their like poles oppositely directed, and 'keepers' made of a magnetic metal (soft iron preferably) are placed in contact with their ends (Fig. 195). A closed magnetic

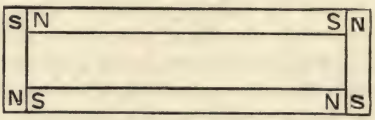


FIG. 195.

circuit is thus formed, for the effect of the magnetism at the poles is annulled by that of the magnetism which is induced in the keepers.]

It appears, therefore, that the form of a magnet must have an effect upon the distribution of magnetisation throughout its interior. Thus, while in a long rod placed parallel to the direction of the force in a uniform field, the magnetisation is sensibly uniform except near the ends, in a short rod the magnetisation is very far from being uniform.

The introduction of a para-magnetic body into a previously uniform field of force disturbs the uniformity of the field. The lines of force, which were originally parallel and equidistant, close in upon the body and become continuous with the lines of induction in its interior.

352. *Permeability and Susceptibility.*—That property of a substance in virtue of which the lines of induction are more or less closely arranged than are the lines of force in the originally undisturbed field is called the *Permeability* of the substance. We have

$$\frac{\mathfrak{B}}{\mathfrak{H}} = \pi \frac{\mathfrak{E}}{\mathfrak{H}} + 1,$$

which may be written in the form

$$\mu = 4\pi k + 1.$$

In this equation, the quantity μ represents the permeability, and k represents the *Susceptibility*. The permeability is therefore the ratio of the induction to the force within the substance of the magnetic body, while the susceptibility is the ratio of the magnetisation to the magnetising force, and is a measure of the readiness of the body to acquire magnetisation.

In a paramagnetic substance, as we have already seen, the lines of induction are more closely arranged than are the lines of force. That is to say, the permeability of such a substance is greater than unity; and therefore the susceptibility is positive. On the other hand, in a diamagnetic substance, the lines of induction are less closely arranged than are the lines of force; and so the permeability

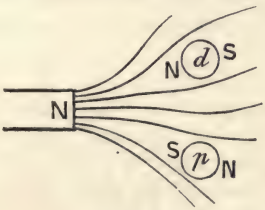


FIG. 196.

is less than unity, and the susceptibility is negative. In a paramagnetic body, north magnetism is manifested at that extremity which faces in the direction in which the external lines of force are drawn; in a diamagnetic body, north magnetism appears at the opposite end. Consequently, while a paramagnetic substance is attracted towards the pole of a magnet, a diamagnetic substance is repelled from it. In Fig. 196, the body marked p is paramagnetic, the body marked d is diamagnetic.

More generally; a paramagnetic substance moves from weak parts to strong parts of a field of force, while a diamagnetic substance moves from strong parts to weak parts.

353. *Residual Magnetism. Retentiveness. Coercive Force.*—We have already seen that some substances, such as steel, retain to a considerable extent their state of magnetisation after the magnetising force is removed. The property in virtue of which this occurs is termed *Retentiveness.*

The magnetism which remains, because of retentiveness, is called *Residual Magnetism.* From its great retentiveness, hard steel is employed in the construction of so-called permanent magnets. The residual magnetism of a long bar of steel is more permanent than is that of a short bar, for the self-demagnetising force (§ 351) has less influence in the former case than it has in the latter.

Retentiveness has very different values in different materials. It is relatively small in good specimens of soft iron.

In order to get rid of residual magnetism in any substance, we must either heat the substance to redness, or employ a reverse magnetising force. It is usual, therefore, to speak of a *Coercive Force* as existent in the material in virtue of which residual magnetism is retained.

354. *Relation connecting Magnetisation and Magnetising Force.*—If the magnitudes of any two of the quantities $\mathfrak{B}$, $\mathfrak{I}$, and $\mathfrak{H}$, are determined in any particular case, the value of the remaining quantity can be calculated from the relation $\mathfrak{B}=4\pi\mathfrak{I}+\mathfrak{H}$. Methods for the determination of each of the three quantities will be described subsequently.

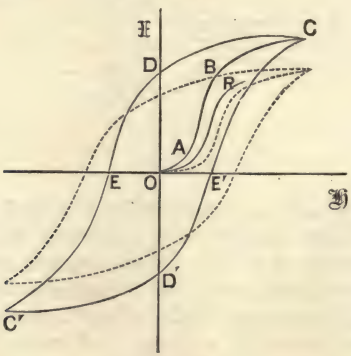


FIG. 197.

Fig. 197 represents the usual course of the variation of intensity with magnetising force. The force is measured along the axis O$\mathfrak{H}$, while the intensity of magnetisation is measured along the axis O$\mathfrak{I}$.

At first, while the force is small, the magnetisation increases very slowly, and at a sensibly uniform rate. Then, as the force is increased, the law becomes $\mathfrak{I}=a\mathfrak{H}+b\mathfrak{H}^2$, a and b being constants. After this, a very slight increase of the magnetising force produces a great change in the magnetisation. With still larger forces, the rate of variation becomes rapidly smaller, and ultimately the magnetisation becomes sensibly constant. These various stages in the process of magnetisation are represented by the parts OA, AB, and BC, of the curve. If the force be now gradually removed, the magnetisation will diminish at a relatively slow rate, until, when the force is entirely removed, a considerable amount of residual magnetisation remains. This is represented by OD.

If a reverse force be now applied, the magnetisation will fall off rapidly in magnitude, and will disappear entirely when the reverse force has a definite value OE. This may be supposed to represent, as Hopkinson suggests, the coercive force.

If the reverse force be now increased until it reaches a value equal to the maximum value of the direct force, if it be then diminished to zero, and if, finally, positive force be reapplied until the original maximum value is attained, the magnetisation will pass through successive values represented by the part EC′D′E′C.

The curve OR represents the residual magnetisation which is left after various magnetising forces have been applied and removed.

The dotted curve represents the change which takes place in the magnetisation when the same substance (say a soft iron wire) is hardened by being stretched beyond its limits of elasticity. The maximum magnetisation is lessened. The residual magnetisation is also lessened, but the coercive force is increased.

Since the magnetisation practically reaches a maximum value when the force is sufficiently great, the substance is then said to be *saturated*. However much the force may be further increased, the intensity remains appreciably constant.

The susceptibility increases from a small value to a maximum which is indicated by the tangent drawn from O to the curve OBC. Thereafter it diminishes to zero as the force increases without limit. The relation $\mu=4\pi k+1$ shows that the permeability also increases from a small value to a maximum (attained at a somewhat greater value of $\mathfrak{H}$ than that at which the maximum susceptibility is reached), after which it gradually diminishes to unity as the force is indefinitely increased.

When soft iron is magnetised (more especially when the force is feeble and the specimen of iron is large) it is found that the magnetisation takes some time to attain its full value corresponding

to the force which is acting. This effect is, by analogy, said to be due to *Magnetic Viscosity*.

355. *Hysterēsis.*—We see from Fig. 197, that the changes of magnetisation tend to lag behind the changes of force which give rise to them. Thus, when the stage C has been reached, a much greater change in the value of the force is requisite in order to effect a given diminution in the magnetisation than was requisite for the production of an equal increase of magnetisation just before the stage C was reached. A similar effect is observable at the point C′. Ewing has called this tendency *Magnetic Hysterēsis*.

As the result of hysterēsis, different values of the magnetisation may correspond to one given value of the magnetic force, and we must therefore limit our definitions of permeability and susceptibility to the case of a substance which is originally unmagnetised, and which is subjected to a force which increases in magnitude continuously from zero upwards.

If E represent the magnetic energy of the magnetised body, the increase of energy which accompanies an increase of intensity of magnetisation $d\mathfrak{I}$ is $\frac{dE}{d\mathfrak{I}}\, d\mathfrak{I}$. Now § 62, $dE/d\mathfrak{I}$ is the force which produces the change $d\mathfrak{I}$: that is, it is the force H. Hence the increment of energy per unit of volume which accompanies the increment of $d\mathfrak{I}$ is $H d\mathfrak{I}$; and therefore (compare § 34) the area CDC′D′C (Fig. 197) represents an amount of energy which has been transformed, per unit of volume, in the given cyclical process. This energy takes the form of heat, and is dissipated. Consequently, rapid reversals of magnetisation will cause the temperature of the magnetised substance to increase markedly; and no amount of lamination, such as is used in transformers or in the armature cores of dynamos for the prevention of heating by induced currents, will prevent this effect.

No dissipation of energy occurs if the cyclical changes in the magnetising force are small, and take place either very rapidly or very slowly. For, in the former case, no time is allowed for a diminution to occur in the amount of lag of the magnetic effect behind the change of force which produces it, whether in the direct or in the reverse part of the cycle; so that the direct changes of magnetisation are exactly reversed when the force is reversed; and, in the latter case, complete time is allowed to prevent noticeable lag from making its appearance, that is to say, the changes in the force take place so slowly that the proper changes of magnetisation can ensue at all stages of the process; and so, again, the reverse changes of magnetisation follow, in the opposite direction, the same course as the

direct changes. In any other case dissipation of energy will be manifested.

It appears, therefore, that the so-called magnetic viscosity tends to produce hysterēsis. But, though this is so, the converse statement that the existence of hysterēsis implies the existence of viscosity is neither necessarily nor actually true.

356. *Effects of Vibration and of Temperature.*—Vibration has a very great influence upon the susceptibility of a magnetised body. This effect is very marked when the magnetising force is small, but is not very noticeable, if at all, when powerful forces are used. It increases the susceptibility of the substance, but diminishes residual

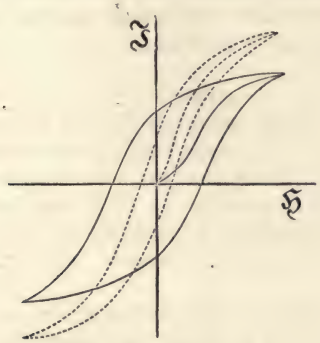


FIG. 198.

magnetism, coercive force, and hysterēsis. These results are exhibited in Fig. 198, in which the full curve represents a cycle performed under the condition of no vibration; while the dotted curve represents the result of an experiment made upon the same substance under similar conditions of force—the substances, however, being tapped after each change in the magnitude of the force.

The temperature of a magnetic substance, too, has a very marked effect upon its susceptibility. In iron, cobalt, and nickel, increase of temperature (from ordinary values) first increases the susceptibility, and afterwards diminishes it, as the magnetising force is continuously increased; and the magnetic properties entirely, and suddenly, vanish when the temperature attains a certain value which is different for each substance, and varies to some extent also from one to another specimen of any substance.

The temperature at which the susceptibility vanishes is called the *Critical Temperature.* It is a temperature at which some fun-

damental change takes place in the physical constitution of the metal. The electric resistance of this metal changes suddenly at this point, as also does its thermo-electric power (§ 328). When the reverse change—from the non-magnetic condition to the magnetic condition—takes place, as hard steel is cooled down from a temperature higher than the critical temperature, a sudden liberation of heat takes place, and the metal glows brightly, although it had previously cooled to dull redness.

In iron, the suddenness with which the magnetisation is lost as the critical temperature is approached, depends very largely upon the value of the magnetising force. When the force is very small, the susceptibility first increases with extreme rapidity to a maximum, and then diminishes with even greater rapidity, as the critical temperature is approached. With higher forces, the variation becomes much less marked.

There is little or no evidence of hysterēsis with regard to the magnetic effects which follow changes of temperature unless the critical temperature be included within the cyclical range. But it at once becomes evident when the range includes the critical temperature; for the temperature at which the magnetic effects reappear as the temperature is reduced, is lower than the critical temperature at which they disappear when the temperature is raised.

This lag of magnetic effect behind the change of temperature which gives rise to it, is abnormally evident in certain alloys of nickel and iron. An alloy containing 25 per cent. of nickel was found by Hopkinson to lose its magnetic properties at a temperature of 580° C., and to remain non-magnetic until its temperature fell somewhat below the Centigrade zero. This fact suggests the idea that non-magnetic manganese steel may become magnetic if its temperature be sufficiently reduced, and that possibly all the non-magnetic metals may act similarly.

357. *Effects of Stress.*—Alteration of the state of stress to which the magnetic metals are subjected, produces considerable alteration of the magnetic qualities of the metals.

Matteuci observed that extension of an iron rod produced an increase of magnetisation, and Villari found that when the field is sufficiently intense, extension causes decrease of magnetisation. This effect is called the 'Villari reversal.'

The various effects of longitudinal and of torsional stress have been very fully investigated by Wiedemann, Sir W. Thomson, and others.

Compression of an iron rod produces effects opposite to those

which are produced by extension. Compression and extension, respectively, of nickel and cobalt rods also produce respectively opposite effects, but there is no Villari reversal in this case; for all values of the magnetising force, extension produces diminution, and compression produces increase, of the magnetisation. The diminution of the residual magnetisation of nickel, under extending stress, is even more evident than is the diminution of induced magnetisation. Hysterēsis, under cyclical variation of load, is much more marked in the case of iron than it is in the case of nickel.

From the above result regarding the effect of extension on the magnetisation of an iron rod in a weak field, we can, by a double application of the principle of stable equilibrium (§ 15), deduce the result that, in weak fields, increase of magnetisation causes increase of length; or, conversely, we can deduce the former result from the latter. Magnetic energy enters the rod from the external medium, and, in part, is transformed into potential energy of molecular configuration within the rod; and this potential energy may be, in turn, transformed into external work as the length of the rod alters. Let us suppose, first, that the length of the rod is not allowed to alter. Increase of magnetisation will then give rise to pressure on the restraining surfaces which prevent alteration of length. Conversely, by the principle of stable equilibrium, decrease of pressure will cause an increase of magnetisation under the given external magnetic force. But, again, increase of pressure upon the restraining surfaces will result in increase of length of the rod if the restraint be removed. Conversely, increase of length will cause a diminution of pressure. We may represent these results by the symbolical expression

$$\begin{array}{ccccc} +\mathrm{M} & \rightsquigarrow & +\mathrm{P} & \rightsquigarrow & +\mathrm{L} \\ & & & & \| \\ +\mathrm{M} & \leftsquigarrow & -\mathrm{P} & \leftsquigarrow & +\mathrm{L} \end{array}$$

where M, P, and L represent respectively magnetisation in weak fields, pressure, and length. Disregarding the intermediate step, the symbols state that increase of the magnetisation of an iron rod in weak fields causes increase of length, and that increase of length of the rod induces an increase of magnetisation. In strong fields the expression would become

$$\begin{array}{ccccc} +\mathrm{M} & \rightsquigarrow & -\mathrm{P} & \rightsquigarrow & -\mathrm{L} \\ & & & & \| \\ +\mathrm{M} & \leftsquigarrow & +\mathrm{P} & \leftsquigarrow & -\mathrm{L} \end{array}$$

where $-\mathrm{P}$ may be translated 'increase of tension.'

[It is important to observe that, when we regard only changes of magnetisation and of pressure (or tension), we are dealing with energy flowing from an external system into the iron; and that, when we regard only changes of pressure (or tension) and of length, we are dealing with energy flowing from the iron rod to another external system; while, when we regard changes of magnetisation and of length alone, we are dealing with energy flowing through the rod from one external system to another, and being transformed through its agency from one form to another.]

Joule proved that no observable change of volume takes place when a rod of iron is magnetised, and therefore that longitudinal magnetisation in weak fields must cause a diminution of the sectional area of the rod. Hence he concluded that if a rod of iron be magnetised circularly, that is, if the lines of magnetisation be circles surrounding the axis of the rod, longitudinal contraction will ensue. He verified this conclusion by experiment.

Torsional strain, also, is accompanied by variations in the magnetic qualities of iron, nickel, and cobalt rods. These effects can, as Sir W. Thomson has shown, be deduced from the known effects of longitudinal stress upon the magnetic qualities. Thus it is known that the susceptibility of iron in weak fields is increased along lines

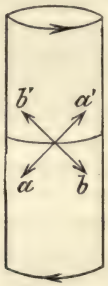


FIG. 199.

of traction, and is decreased along lines of compression. But, when a circular rod of iron is twisted in the manner which is indicated by the arrows in Fig. 199, all lines such as $a\ a'$ suffer traction, while all lines such as $b\ b'$ suffer compression. Hence the susceptibility is increased along $a\ a'$ and is diminished along $b\ b'$. The effect of this is practically to produce two components of magnetisation—one longitudinal, the other circular—when the twist is sufficiently great.

Hence torsion in weak fields diminishes the longitudinal susceptibility of iron.

Conversely, a circularly magnetised iron rod, when twisted, becomes longitudinally magnetised. It is easy to deduce the corresponding effects in nickel and cobalt. [No reversal of the direction of longitudinal magnetisation takes place in iron, however strong the circular magnetisation may be. Ewing explains this by the fact that the intensity of magnetisation in the direction of the line of traction, or of compression, never reaches the point at which the Villari reversal occurs.]

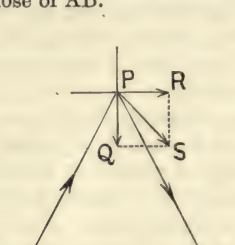
Since torsional stress produces circular magnetisation in a longitudinally magnetised rod, and since it also gives rise to longitudinal magnetisation in a circularly magnetised rod, we might expect that the superposition of longitudinal and circular magnetisations would cause torsional strain. This effect was discovered experimentally by Wiedemann in the case of iron. The twist in weak fields takes place in such a direction as to be completely explainable by the increase of length which occurs in the direction of resultant magnetisation. Knott has shown that the twist occurs in the reverse way in nickel—a result which he expected to find, since nickel contracts in the direction of magnetisation.

We may observe here that the twisting of a magnetised rod, or the magnetisation of a twisted rod, gives rise to a transient electric current in the magnetised material. This effect will be considered in next chapter.

358. *Magnetometric Measurements.*—The *magnetometer* consists essentially of a small magnet, which is suspended by a long fine fibre, whose co-efficient of torsion is negligeable in most cases, and which is free to turn about that fibre as an axis. A small mirror is usually attached to the magnet, so that, by means of a reflected beam of light, very small angular motions of the magnet may be made evident. This apparatus is placed in a uniform field of force of known intensity, say the earth's field (§ 359). The magnet then places its length in the direction of the force in the given field.

Let the magnet be placed at P (Fig. 200), and let the direction of the controlling force be PQ. Let AB be a bar magnet, the intensity of magnetisation of which we have to determine, and let the points A and B represent the position of its poles. Place it symmetrically with regard to PQ in the position which is indicated in the figure. Let I be its (unknown) intensity of magnetisation, while a is its sectional area. Then Ia is the strength of its pole. The effect of the north pole, A, at P is in the direction AP, and is equal to

Ia/AP^2. Similarly, the directive force of the south pole, B, at P is in the direction PB, and is equal to $-Ia/PB^2$. If we represent these forces by AP and PB respectively, it is obvious that the resultant effect of the two is represented on the same scale by AB. The magnitude of the resultant is therefore $IaAB/AP^3$, and acts so as to place the little magnet at P parallel to AB, with its poles facing oppositely to those of AB.



F G. 200.

Now let PR represent this force on the same scale that PQ represents the force of the external field. PS is the resultant of these forces, and the little magnet sets itself in the direction PS, making an angle θ with PQ, such that $\tan\theta = PR/PQ$. If F be the intensity of the external force, this gives

$$\frac{IaAB}{AP^3} = F\tan\theta, \ldots\ldots\ldots (1)$$

from which we can calculate I.

In order to find the value of the force F, if it be unknown, we may set the magnet AB oscillating under the action of F alone. The time, T, of a small oscillation is given by the equation

$$\frac{4\pi^2}{T^2} = \frac{FIaAB}{K},$$

where K is the moment of inertia of the magnet about its axis of suspension, and AB is its length. For if the direction of the force F be denoted by the arrow (Fig. 201), and if *ns* represent the magnet inclined at an angle θ to the direction of the force, FS in θ is the force which is acting perpendicularly to the length of the magnet, and which tends to turn it around its axis of suspension. This force acts at each end of the magnet so as to produce rotation

in the positive direction, and the turning moment is therefore $F \sin\theta\, I a AB$. When the angle is small, this becomes $F\theta I a AB$. The angular acceleration is $\ddot{\theta}$ (§§ 42, 45), and the momentum which is porduced per unit of time is $m\ddot{\theta}R$, where m is the mass of the magnet

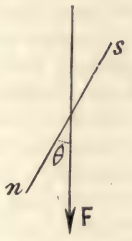


FIG. 201.

and R is its radius of gyration (§ 70). Hence the moment of momentum which is produced per unit of time is $m\ddot{\theta}R^2 = K\ddot{\theta}$, where K is the moment of inertia (§ 70). We therefore have

$$K\ddot{\theta} = -F\theta I a AB,$$

the minus sign being used, since the angular acceleration is negative. Now every quantity in this equation is constant, with the exception of θ; and so the equation expresses the fact that the angular acceleration is negative and is proportional to the displacement. The small oscillations of the magnet, therefore, obey the simple harmonic law, and the angular position of the magnet is given (§ 51) by the equation

$$\theta = P\cos\left(\sqrt{\frac{FIaAB}{K}} \cdot t + Q\right),$$

where P and Q are constants and t is the time; whence T being the periodic time, we get (§ 51)

$$\frac{4\pi^2}{T^2} = \frac{FIaAB}{K} \quad \ldots\ldots\ldots (2).$$

By elimination between the equations (1) and (2), we can find the values of F and of $IaAB$ (which is the magnetic moment of the magnet). Also, a and AB being known, we can obtain the value of I; and hence, if we know the intensity of the magnetising force, we can calculate the susceptibility and the permeability of the substance.

Equation (2) shows that the intensities of two fields are inversely

proportional to the squares of the periods of the small oscillations of a magnet of known magnetic moment, which is suspended first in one field and then in the other.

The ballistic method of making magnetic measurements will be discussed in § 369.

359. *Terrestrial Magnetism.*—The earth exerts a magnetic action in virtue of which compass needles point in a northerly direction. The angular distance between the line along which a compass needle points and the geographical meridian is called the magnetic *declination* or *variation*. The magnetic needle, if it were carefully supported on an axis which passes through its centre of inertia, and which is perpendicular to the magnetic meridian, would, in Britain, place its magnetic axis in a direction which is inclined to the horizon—the north end pointing downwards at a considerable angle. This angle is called the *Magnetic Dip*.

The declination and the dip vary considerably from one part of the earth's surface to another. In some regions the declination is easterly, in others it is westerly. The line on the earth's surface, at all points of which the declination is zero, is called the *Magnetic Equator*. It does not coincide with the geographical equator, and is not a great circle. That point on the surface at which the north pole of a magnet points vertically downwards is called the *North Magnetic Pole* of the earth, and that point at which the south pole of a magnet points vertically downwards is called the *South Magnetic Pole* of the earth. These poles do not coincide with the geographical poles of the earth, neither do they lie at opposite extremities of a diameter. [Observe that the magnetism which we may suppose to be collected at the *north* pole of the earth must be *south* magnetism, *i.e.*, it must be of the same kind as that which appears at the south-pointing pole of a magnet. Similarly, the magnetism which is manifested at the south magnetic pole of the earth must be north magnetism.]

The earth's magnetic force is in a constant state of variation. It changes with the hour of the day and the time of the year; and it depends also upon the position of the moon. Yet these variations do not appear to be due to any *direct* action of the sun or the moon.

Sudden disturbances sometimes take place in addition to these more regular variations. A period of maximum disturbance occurs every eleven years, and coincides with the period of maximum sun-spot disturbance.

A slow secular change of the position of the magnetic poles is also in progress.

360. *Theories of Magnetism.*—At one time magnetic phenomena were explained by the assumption of the existence of two imponderable fluids, one of which constituted north magnetism, while the other constituted south magnetism. In Poisson's elaboration of this theory a magnetic body was supposed to be made up of spheres of infinite permeability, uniformly distributed in an absolutely non-permeable fluid. This made the problem of magnetic induction identical with that of electric induction in a non-conducting medium, throughout which perfectly conducting insulated spheres were uniformly distributed. Among other objections to this theory is the fatal one pointed out by Maxwell, that the permeability of iron is too great to be accounted for even if the spheres were packed in the closest possible arrangement.

In modern theories the molecules are supposed to be little magnets. In an unmagnetised body the magnetic molecules have their axes distributed, on the whole, uniformly in all directions; and the substance becomes magnetised when the axes of its molecules get, on the whole, a definite set in one direction. Saturation will take place when all the molecules have set their axes in the direction of the magnetising force.

The fact that the slightest force does not produce saturation shows that displacement of the molecules must be resisted by some force. Weber assumed that each molecule is acted upon by a constant force in the original direction of its axis, which tends to prevent its orientation from that direction, and tends to make it resume its original direction when it is displaced from it. It follows from this assumption that the curve of magnetisation (Fig. 197) should at first be a straight line, that it should afterwards become concave to the axis along which the force is measured, and that it should ultimately approach an asymptote parallel to that axis. This does not agree with experiment.

Maxwell improved this hypothesis by the additional assumption that a molecule could return to its original position if it were turned through an angle of less than a certain finite magnitude, and that if it were displaced through an angle greater than this, it would retain, after removal of the force, a displacement equal to the excess of its total displacement over this quantity. This form of the theory leads to a magnetisation curve similar to that given by Weber's unmodified theory, and it indicates that the curve of residual magnetisation starts from a point on the force-axis at a finite distance from the origin, is always concave to that axis, and approaches a parallel asymptote. These results also are incorrect.

Ewing, following a limit by Maxwell, regards each molecule as

subject only to the mutual action of the entire system of surrounding molecules. He has constructed a model of such a system by means of a number of pivoted magnets, which are arranged in parallel rows. So long as no external magnetic force acts, the magnets arrange themselves in positions of stable equilibrium under their mutual forces, some of them pointing in one direction, some in another. This illustrates the condition of non-magnetised steel. If only a feeble uniform magnetic force acts, each magnet is slightly turned from its first position, which it reassumes when the force is removed. This illustrates the first stage in the process of magnetisation. A somewhat stronger force causes instability in the originally less stable groups of the magnets, and the magnets which compose these groups swing round into a new stable position. As the external force is increased still further, more and more groups break up, until all have taken the new position of equilibrium under their own mutual forces and the external directive force. This illustrates the second stage of magnetisation, in which the ratio of magnetisation to magnetising force increases with great rapidity. The third stage, in which this ratio is practically constant, is exemplified by the fact that an infinite force is now needed to make the magnets point exactly in the direction of the external lines of force. If the external force be now removed, a considerable proportion of the magnets retain their final positions of equilibrium—in other words, magnetic retentiveness is exhibited.

This model can also show the effects of strain on the magnetic properties. For this purpose the magnets are placed on a sheet of indiarubber. If the indiarubber is stretched the magnets are separated out from one another in one direction, and are brought nearer to each other in a direction at right angles to the former. The magnetic susceptibility is increased or is diminished, according as the stability of the magnets is diminished or is increased by the alteration of relative position. Similarly, the increase of the susceptibility of iron with rise of temperature is explained by the diminution of mutual magnetic influence which results from increased distance. Professor Ewing suggests that the total loss of magnetisation which occurs at a high temperature is due to a continuous whirling motion of the magnetic molecules. He suggests, also, that the dissipation of energy, which occurs when hysteresis is exhibited, is due to the induced electric currents (§ 342), which are caused by angular motions of the magnetic molecules.

CHAPTER XXXI.

ELECTROMAGNETISM, ETC.

361. *Oersted's Discovery.*—Oersted found that a magnetic needle always tends to place its length in a direction at right angles to a plane which, passing through its centre, contains a linear circuit, through which an electric current is flowing. The direction in which the north pole points depends upon the direction in which the current flows. The north pole always tends to move round the linear circuit in a direction which is related to that of the current in the way in which the rotation of a right-handed screw is related to its linear motion.

After this fact was discovered, it was surmised that the converse phenomenon might also be found to exist—that motion of the linear circuit, through which the current flows, would take place if the magnet were fixed while the circuit was free to move. This was verified experimentally by Ampère.

And, further, since a magnet can act thus upon two neighbouring circuits, through which electric currents flow, it was supposed that mutual action might be found to exist between these circuits themselves if the magnet were removed. Ampère proved the existence of this action also.

362. *Magnetic Action of Closed Electric Circuits.*—The above statement regarding Oersted's discovery shows that a linear electric current is surrounded by circular magnetic lines of force. The researches of Ampère and of Weber have enabled us to find the distribution of electric force in the neighbourhood of *any* conducting circuit.

It is found that a small plane closed circuit produces the same magnetic action as a small magnet, placed at some point inside the circuit with its length perpendicular to the plane, and having the direction of its axis (measured from south pole to north pole) related to the direction of the circulation of the current according to the rule of right-handed screwing motion, while its magnetic

moment is equal to the area of the circuit multiplied by the strength of the current. [The word *small* means that the point at which the action is determined is very far off in comparison with the dimensions of the circuit.]

It is of no consequence in what part of the circuit the equivalent magnet is placed; and so we may assume that it is an extremely thin magnetic shell, which fills the entire circuit, and is possessed of a magnetic intensity which is numerically equal to the strength of the electric current divided by the thickness of the shell. For, if I, a, and t represent respectively the magnetic intensity, the area, and the thickness of the shell, Iat is the moment of the shell, and, therefore, $ia = \mathrm{I}at$, or $i = \mathrm{I}t$, where i is the strength of the current, and It is called the strength of the shell.

Now, let any finite circuit PQRS (Fig. 202) be filled with a network of infinitely small conducting meshes of any shape. Let a current i circulate in the circuit in the positive direction. We may assume that an equal current flows similarly in each of the meshes, such as $pqrs$, for the currents in each common side of two adjacent

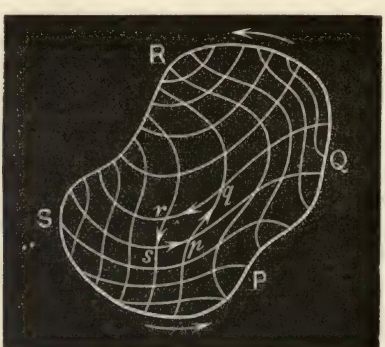


Fig. 202.

meshes exactly neutralise each other. But, as above, the magnetic effect of the current in each mesh may be supposed to be due to a magnetic shell of strength i. And hence we see that the magnetic action of the circuit PQRS at external points is equivalent to that of a magnetic shell of strength i, *and of any form*, which completely fills the circuit.

It is easy to find a simple expression for the magnetic potential of the shell. For, if dS be an element of its surface, we may replace the part of the shell which corresponds to dS by a small magnet,

the strength of whose pole is $id\mathrm{S}/t$. The force with which the north pole n (Fig. 203) acts on a unit north pole, placed at a point P, is $id\mathrm{S}/tr^2$, where r is the distance from P to n. The potential at P

FIG. 203.

due to n is, therefore, $id\mathrm{S}/tr$. Similarly, the potential of s at P is $-id\mathrm{S}/tr'$, where r' is the distance from P to s. The total potential is, therefore,

$$\mathrm{V}=\frac{id\mathrm{S}}{t}\left(\frac{1}{r}-\frac{1}{r'}\right)=\frac{id\mathrm{S}(r'-r)}{tr^2}$$

since r' is practically equal to r. But $r'-r=t\cos\theta$, where θ is the angle between the axis of the magnet and the line joining P to its centre, and t is the length of the magnet (equal to the thickness of the shell). Hence

$$\mathrm{V}=\frac{id\mathrm{S}\cos\theta}{r^2}.$$

Now, since the element of the surface of the shell is at right angles to the axis of the magnet, $d\mathrm{S}\cos\theta$ represents the resolved part of the element normal to r, and $d\mathrm{S}\cos\theta/r^2$ is the elementary solid angle which the surface subtends at P. We may, therefore, write

$$\mathrm{V}=id\omega$$

where $d\omega$ represents this elementary angle. And, in order to find the total potential V′ at P, due to the whole shell, we have merely to sum all the quantities, such as V, for the whole surface. Hence

$$\mathrm{V}'=i\omega,$$

that is, the potential at any point external to the magnetic shell is equal to the product of the intensity of the current into the solid angle which the shell subtends at the given point.

It follows that the work which is done upon a unit north pole by the magnetic forces due to the current is $i(\omega_1-\omega_2)$, if the pole passes from a place where ω has the value ω to a place where it has the value ω_2. In particular, if the pole completely describes a closed path which does not pass through the interior of the circuit in which the current flows, the work is zero. If the pole passes

by an external path from a point P, which is infinitely close to one side of the shell, to a point P′, which is infinitely near to P on the opposite side of the shell, ω changes by an amount which is infinitely nearly equal to 4π, and the work which is done by the magnetic forces is practically $4\pi i$. But *any* shell of the proper moment, which is bounded by the circuit, produces the same magnetic effect at distant points as the current produces; and so we may now replace the shell, whose action we are considering, by another shell of the same moment, which is everywhere finitely distant from P and P′, and which is bounded by the same circuit. The forces due to this shell do infinitely little work upon the unit north pole when it passes from P′ to P, and hence the work which is done upon the unit pole in a single complete passage round a closed path which passes through the circuit is $4\pi i$. In n such passages the work is $4\pi in$.

363. *Electrodynamic Action on an Electric Circuit.*—Let the intensity of magnetisation of a shell, which produces the same magnetic action as a given circuit, be I; and let the shell be placed in a field of force the intensity of which is F. We may choose the shell so that each element of its surface is at right angles to the direction of the force in its immediate neighbourhood. Consider a small portion ds of the north face of the surface. The total force acting on this part is $+\mathrm{F}\mathrm{I}ds$, the plus sign being used since we consider it to be positive when it acts in the direction of the outward normal to the north face. Similarly the force which acts on the corresponding part of the south face is $-\mathrm{F}\mathrm{I}d\mathrm{S}$; and therefore the potential energy of this portion of the shell in the given field is $-\mathrm{F}\mathrm{I}d\mathrm{S}t$, where t is the thickness of the shell. Now $\mathrm{I}t$, the strength of the shell is equal to i, the strength of the current; and so the potential energy is $-\mathrm{F}id\mathrm{S}$. Consequently, the total potential energy of the shell is

$$-i\mathrm{N}$$

where N is the whole force acting on the shell, that is, the number of lines of force which pass through the circuit whose action is represented by that of the shell.

When a force F produces a change df in one of the quantities which determine the position of the circuit, the work which is expended is $\mathrm{F}df$, and the corresponding change in the potential energy of the circuit is $id\mathrm{N}$. Hence

$$\mathrm{F} = i\frac{d\mathrm{N}}{df}.$$

And we see that the force F tends to produce or to oppose the change df according as that change is accompanied by an increase,

or by a decrease, of the number of lines of force which pass through the circuit in the positive direction.

In particular, if a part of the circuit be moveable, the electromagnetic forces which act upon the circuit will tend to produce such a displacement of the moveable part as will cause an increase of the number of lines of force which pass through the circuit.

364. *Case of Linear Circuits.*—We have already seen that a linear circuit carrying an electric current is surrounded by circular lines of force, the direction of which is related to that of the current in the same way as the rotation of a right-handed screw is related to its linear motion.

Let A B (Fig. 204) represent part of a fixed linear circuit through which a current flows from A to B, and let *ab* be a portion of a moveable parallel circuit through which a current flows from *a* to *b*. We may assume that the circuit *ab* is completed by way of *p*. The lines of force due to AB pass through *abp* in the positive direction, and a displacement of *ab* towards AB would increase the area *abp*,

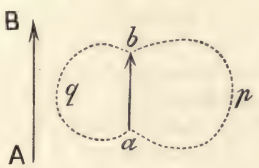


FIG. 204.

and so would cause an increase in the number of positively drawn lines of force. Hence the electrodynamic action between the circuits is such as to cause their mutual approach. [We would arrive at the same result by the supposition that the circuit *ab* is completed by way of *q*. For the lines of force due to AB pass through the electric circuit *abq* in the negative direction. Hence *ab* will move so as to diminish the area of *abq*, that is, so as to diminish the number of negatively drawn lines of force which pass through it.] Similarly, mutual repulsion will ensue if the currents are oppositely directed.

Next let the circuits AB and *ab* be inclined to each other, and let OO′ (Fig. 205) be the shortest line between them. Let us suppose that, while AB is fixed, *ab* is free to turn around OO′ as an axis. Complete the circuit *ab* by way of *p*. No lines of force due to AB will pass through *abp* when AB and *ab* are mutually perpendicular. On the other hand, the number of lines which pass through it in the posi-

tive direction is a maximum when the currents in AB and *ab* are parallel and similarly directed. The electrodynamic action on *ab* is

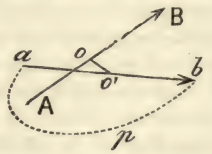


FIG. 205.

therefore such as to cause it to place itself parallel to AB. If either current be reversed the moveable circuit will turn so that *ba* is co-directional with AB.

365. *Circular Circuits. Solenoids. Ampère's Hypothesis regarding Magnetism.*—A circular circuit, of radius r, which is traversed by a current i, may be replaced by its equivalent magnetic shell of intensity $\pm i/t$ (§ 362). So also, a second circular circuit, which is traversed by a current i', may be replaced by a shell of intensity $\pm i'/t'$. These shells will turn so as to place their oppositely magnetised forces parallel, and will then exhibit mutual attraction. Hence the electric circuits will tend to turn so that the currents in each are parallel, and will then exhibit mutual attraction. But if, while the circuits are in this position, one of the currents be reversed, mutual repulsion will be exhibited.

[It is an easy matter to calculate the force which such a circuit carrying a current i, exerts at its centre. Let the centre be at o

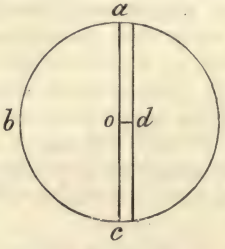


FIG. 206.

(Fig. 206) and let *a*, *c*, be the points in which the circuit cuts the plane of the paper—the plane of the circuit being supposed to be perpendicular to that plane. Imagine the circuit to be replaced by an equivalent hemispherical magnetic shell (§ 362) *abc*. The work

which is done in displacing a unit pole from O through a small distance $Od=\tau$ is $i(\omega-\omega')$, where ω and ω' are respectively the angles subtended at O and d by the shell abc. If r be the radius, the value of ω is 2π, and the value of ω' is practically $(2\pi r^2-2\pi r\tau)/r^2$; so that the work is $2\pi i\tau/r$. Hence the force, which is practically uniform when τ is sufficiently small, is $2\pi i/r$; so that, when r is unity, unit length of the current exerts a force i at the centre.]

These phenomena can be readily exhibited by means of two small floating cells, each of which consists of a test tube containing dilute sulphuric acid into which dip zinc and copper plates connected externally by a circular copper wire. The test tubes are inserted in pieces of cork, and are floated on the surface of water.

A wire which is bent into a cylindrical helix in the manner indicated in Fig. 207, is called a *Solenoid*. If it be freely suspended on pivots, and be traversed by a current, it will act like a magnet under the action of the earth's force or of other magnets or solenoids. If the number of turns, n, per unit of length is large, we may

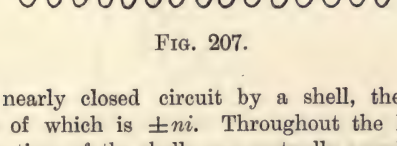


FIG. 207.

replace each nearly closed circuit by a shell, the intensity of magnetisation of which is $\pm ni$. Throughout the length of the solenoid, the actions of the shells are mutually annulled, except at the ends, where quantities of magnetism $\pm nia$ are found, a being the area of the shells. At points which are far distant in comparison with the radius of the solenoid, the action is therefore equivalent to that of a magnet of moment $nial$, where l is the length of the solenoid.

In the interior of a very long solenoid, which contains n turns per unit of its length, and through which a current i circulates, the total force is equal to $4\pi nia$, where a is the area of a transverse section. For the thickness of each shell equivalent to a turn of the wire is $1/n$, and therefore the surface density of the distribution of magnetism on its two faces is $+ni$ and $-ni$ respectively. Hence the force at any point within each shell, and therefore throughout the interior of the solenoid, is $4\pi ni$. Thus a solenoid may be used for the production of a very intense magnetic field; and this

furnishes one of the most convenient ways of temporarily, or permanently, magnetising a magnetic substance.

These properties of circular circuits and of solenoids led Ampère to suggest that the molecules of magnetic substances may exhibit their magnetic properties in virtue of electric currents which circulate within them in closed circuits.

366. *Continuous Rotation under Electromagnetic Force. Electric Motors.*—We have already seen that no work is, on the whole, done upon a magnetic pole which describes a closed path in a field of force due to an electric circuit, provided that the path does not pass through the interior of the circuit. But it is also true that no work will on the whole be done upon a magnetised body which completely describes a closed path passing through the interior of the circuit; for the body is composed of excessively small magnetised molecules, and the total amount of work which is expended upon each molecule in the process is zero, since its north and south poles are of equal strength.

But work will be expended on the whole if motion of part of the circuit takes place, without interruption of the current, under the action of external magnetic force. As an example, let us consider a horizontal circular conductor AB (Fig. 208). A current which

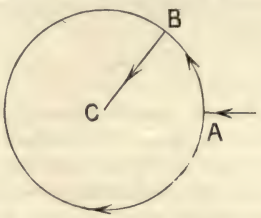


FIG. 208.

enters this circuit at A, will divide into two parts, which reunite at B and flow through the conductor BC to the point C, which is connected with the negative pole of the battery to the positive pole of which the point A is joined. The lines of force, due to the earth's magnetic action, pass downwards through the circuit. In the region ABC, their direction is related to that of the current according to the law of left-handed screwing motion: in the region to the other side of BC, their due action is related to that of the current according to the law of right-handed screwing motion. Hence the electrodynamic action upon BC will cause it to rotate in the direc-

tion of the hands of a watch provided that it is pivoted at C, and has a sliding contact at B.

If AB were a circular conducting disc, pivoted at C, and having a sliding contact at its circumference, so that an electric current flowed radially inwards, and if lines of force (whether due to the earth or to external magnets) passed through it as above, continuous rotation of the disc in the direction of the hands of a watch would ensue. This arrangement is known as Barlow's wheel.

Conversely, continuous motion of the magnet may take place if the circuit be fixed while the magnet is free to move and the direction of the current is reversed whenever the magnet passes from one side of the circuit to the other. Indeed, it is easy to arrange the conditions in such a way that continuous rotation of the magnet will take place without periodic reversal of the direction of the current. For example, let *ns* and *n's'* (Fig. 209) be two magnets,

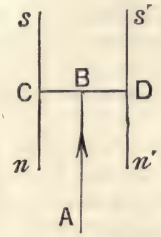


FIG. 209.

which are connected together by the cross piece CD, and which, being pivoted at B, are free to rotate about AB; and let a current flow continuously along AB. The direction of the circular lines of force which surround AB is related to that of the current according to the law of right-handed screwing motion; and therefore the poles *n* and *n'* rotate around AB in that direction. And it follows that a delicately pivoted magnet, along which a current flows, will be similarly set in rotation, for it may be supposed to consist of a number of magnets grouped around its axis.

These principles of electrodynamic action are applied practically in the construction of electro-magnetic machines or motors for the transformation of electric energy into mechanical work. The electric circuits may be fixed while the magnets rotate; or, preferably, the magnets may be fixed while the circuits rotate.

367. *Electromagnetic Induction. The 'Dynamo.'*—We have seen that the increase of the potential energy of an electric circuit through

which a current i flows is $-id\mathrm{N}$, where $d\mathrm{N}$ is the increase of the number of the lines of force which pass through the circuit in the positive direction. Conversely, the work which is done upon the circuit by the electromagnetic forces during a process in which the number of lines of force which pass through the circuit increases by the amount $d\mathrm{N}$ is $td\mathrm{N}$. If the conditions are such that this work can be transformed into electric energy in the circuit, a reverse electromotive force must be produced which opposes the passage of the current i. Now this electric energy is developed to the amount $i\mathrm{E}$ per unit of time, where E is the reverse electromotive force (§ 342). Therefore, since $d\mathrm{N}/dt$ is the increase of N per unit of time, we get

$$\frac{d\mathrm{N}}{dt} = \mathrm{E}\,; \ldots\ldots\ldots (1),$$

that is, *a reverse electromotive force acts around the circuit which at any instant is measured by the rate of increase of the number of lines of force which pass through the circuit.* [It must be remembered that a reverse electromotive force is one which tends to produce a current in the circuit in a direction which is related to the direction of the lines of force according to the law of left-handed screwing motion.]

The phenomenon, whose existence we have here assumed, was discovered experimentally by Faraday. He found that if, from whatever cause, the number of lines of force passing through a circuit is increased, a reverse electromotive force will act round the circuit, and will produce a reverse current; and that, if the number of lines be decreased, a direct force will act and will produce a direct current. The currents so produced are called *induced currents.* They only last so long as there is a variation of the number of lines of force in progress. In particular, they may be produced by the electromagnetic action which is due to varying currents in other fixed circuits, or to the motion of other circuits which carry steady currents; or they may be due to the action of moving magnets.

For example, if a current be started in one direction in a linear conductor, a transient current will flow in the opposite direction in a parallel conductor; and, if the direct current in the former be stopped, a transient direct current will flow in the latter. Phenomena such as these are called phenomena of *mutual induction.*

But it is important to observe that the number of lines of force which pass through a closed circuit depends upon the current which is flowing through that circuit as well as upon external currents. These lines of force pass in the positive direction through the circuit, and so any increase in their number, due to an increase in the

strength of the current, causes a reverse electromotive force in the circuit which prevents the direct current from instantly attaining its full strength under the action of a suddenly introduced electromotive force, or from instantly falling to zero when the force is removed. This phenomenon is called *self-induction*, and was investigated experimentally by Faraday. It can only be prevented by arranging the circuit in such a way that its total area is zero. For example, if a plane circuit be crossed upon itself in a figure-of-eight shape, so that the areas of the two loops are equal, no self-induction will occur, for the lines of force which pass through each loop are equal in number, but are oppositely directed.

The number of lines of force which traverse one circuit because of the electromagnetic action of another circuit through which a current j is flowing is

$$\mathrm{N}=j\mathrm{M}, \quad \ldots\ldots\ldots\ldots (2)$$

where M is a quantity which depends upon the form and mutual position of the two circuits (§ 362), and is called the *co-efficient of mutual induction* of the two circuits.

Let a constant electromotive force J act in the circuit through which the current j is flowing, and let an electromotive force I act in another circuit, through which a current i flows, and whose co-efficient of mutual induction with regard to the former circuit is M. Also let the resistance of the former circuit be S, while that of the latter is R; and let the co-efficient of self-induction of the former be Q; while that of the latter is P. We get

$$\mathrm{I}=\frac{d(j\mathrm{M})}{dt}+\frac{(di\mathrm{P})}{dt}+i\mathrm{R}, \quad \ldots\ldots (3)$$

$$\mathrm{J}=\frac{d(i\mathrm{M})}{dt}+\frac{d(j\mathrm{Q})}{dt}+j\mathrm{S}. \quad \ldots\ldots (4).$$

If the two circuits are fixed, these equations become respectively

$$\mathrm{I}=\mathrm{M}\frac{dj}{dt}+\mathrm{P}\frac{di}{dt}+\mathrm{R}i, \quad \ldots\ldots\ldots (5)$$

$$\mathrm{J}=\mathrm{M}\frac{di}{dt}+\mathrm{Q}\frac{dj}{dt}+\mathrm{S}j \quad \ldots\ldots\ldots (6).$$

The first term on the right hand side of these equations represents the reverse electromotive force due to mutual induction; the second represents the reverse force due to self-induction; and the third represents (§ 336) the part of the electromotive force which maintains the current against the resistance of the circuit.

If the circuit in which J acts be entirely removed, the equation which applies to the other circuit becomes

$$I = P\frac{di}{dt} + Ri.$$

Let us suppose that the force I is maintained until the current attains a steady value i_o, after which I is withdrawn. We then have

$$P\frac{di}{dt} + Ri = o,$$

which gives (§ 38)

$$i = i_o \varepsilon^{-\frac{R}{P}t}.$$

This shows that the intensity of the current diminishes in geometrical progression as the time increases in arithmetical progression, and that it (theoretically) takes an infinite time to reach zero intensity. Practically, the condition of zero intensity is in most cases attained in a small fraction of a second. Similarly, we find that the equation

$$i = i_o\left(1 - \varepsilon^{-\frac{R}{P}t}\right)$$

represents the relation between the current i, at a time t after the electromotive force I begins to act, and the steady current i_o.

The introduction of iron cores into the circuits greatly increases the self and mutual induction, because of the great permeability of iron. This is the essential principle of the Ruhmkorff coil, which consists of a coil of stout, insulated copper wire wound round an iron core, and surrounded by another coil of very fine, well-insulated copper wire. The inner coil is called the *primary* coil, the outer is called the *secondary* coil. The former has very small resistance, while the latter has very high resistance. The core is composed of a number of fine iron wires for the purpose of preventing the induction of currents within it, for these currents act so as to oppose the direct induction. By this means feeble electromotive forces in the primary circuit may give rise to very high electromotive forces in the secondary circuit.

In the modern 'dynamo' coils of wire with iron cores are caused to move rapidly between the poles of a powerful electromagnet. Induced currents are thus produced in the coils, and may be used for purposes of electric lighting, etc. We cannot here enter into a discussion of the various ingenious details of construction which are adopted in these machines in order to secure high efficiency, or to

adapt them for the performance of different duties. The subject has now a complete literature of its own.

Arago found that a magnet, which is pivoted above a horizontal copper disc, will be set into rotation if the disc be rotated. Faraday explained this by the electromagnetic action of the currents which are induced in the disc. If radial slits be cut in the disc, the action will greatly cease; for the induction of currents is prevented except on a small scale.

Currents are induced in the body of a magnet itself whenever its state of magnetisation varies. Thus, since an electric current flowing along a rod is surrounded by closed lines of magnetic force, conversely, any change in the circular magnetisation of a rod will cause the flow of a transient current along the rod. This may readily be made manifest by connecting the ends of a twisted iron rod to the terminals of a galvanometer (§ 369), and suddenly magnetising the rod longitudinally. Since the rod is twisted, longitudinal magnetisation cannot occur without circular magnetisation (§ 357), and so a transient longitudinal current occurs, and is made manifest by the galvanometer. The same effect is produced if a longitudinally magnetised rod be suddenly twisted.

368. *Electrokinetic Energy.*—Consider again the two electric circuits dealt with in last section. Let them move so that M increases by the amount $d\mathrm{M}$, and let the motion be so slow that the currents i and j are sensibly constant. The rate at which heat is developed per unit of time in the two circuits is

$$\mathrm{H}=\mathrm{R}i^2+\mathrm{S}j^2,$$

and the rate at which energy is supplied per unit of time in maintaining the electromotive forces I and J constant is

$$\mathrm{E}=\mathrm{I}i+\mathrm{J}j.$$

Hence
$$\mathrm{E}-\mathrm{H}=i(\mathrm{I}-\mathrm{R}i)+j(\mathrm{J}-\mathrm{S}j);$$

which becomes
$$\mathrm{E}-\mathrm{H}=2ij\frac{d\mathrm{M}}{dt},$$

since, in (3) and (4) above, if i, j, P, and Q are constant, we get

$$\mathrm{I}-\mathrm{R}i=j\frac{d\mathrm{M}}{dt},$$

and
$$\mathrm{J}-\mathrm{S}j=i\frac{d\mathrm{M}}{dt}.$$

But by (1) and (2) we see that

$$ij\frac{d\mathrm{M}}{dt}$$

represents the rate at which work is done in the circuit by electrodynamic action, and thus the equation shows that, under the given conditions, an amount of energy must be drawn from the source in a given time which exceeds that developed in the circuits in the form of heat by twice the amount of work which is simultaneously performed by the electromagnetic forces. This excess is called the *electrokinetic* energy of the system. In Maxwell's theory it is supposed to reside in the medium which surrounds the circuits. It is transformed into heat, etc., whenever the circuits are broken; for the rupture of the circuits is, under these conditions, attended by a spark of more than usual intensity.

A similar result can be deduced from (5) and (6) when the circuits are fixed and i and j vary.

369. *The Galvanometer. The Ballistic Method.*—A galvanometer is an instrument by means of which the intensity of an electric current is measured through the magnetic effect which the current produces. It (as already stated) usually consists of coils of wire in the interior of which a small magnet is freely suspended. In its normal position the magnet has, under the action of an external force, its length perpendicular to the axis of the coil; and the intensity of a current is proportional to the tangent of the angle through which the magnet is deflected from its normal position when a current passes through the coil, for the current produces in the interior of the coil a practically uniform magnetic field, whose intensity is proportional to the current-strength and whose direction is parallel to the axis (§ 365), so that an equation of the form (1), § 358, applies.

In other forms of the instrument the coil is freely suspended in a constant magnetic field; and the name *Electrodynamometer* is given to an instrument in which both of the magnetic fields are produced by the current, which flows simultaneously through two coils, one of which is fixed, while the other is freely suspended in its interior or swings freely around it. The indications of the latter instrument are proportional to the square of the current-strength.

In the Ballistic Galvanometer the suspended portion has great moment of inertia, and, consequently, has a long period of vibration (§ 131). When a transient current, whose duration is very small in comparison with the periodic time, passes through this instrument, the total quantity of electricity which passes is proportional to the sine of half the angle of deflection. Hence the instrument may be applied to the investigation of magnetic properties. For example, if a coil of wire connected with the galvanometer be

wound on an iron bar the intensity of magnetisation of which is varied from time to time, the transient current which follows each variation of intensity produces a deflection in terms of which the total quantity of electricity which passes can be calculated. Now, if dN be the change of induction through the coil, we get by (1), § 367,

$$dN = E dt,$$

where E is the electromotive force, which is equal to Ri, if i is the intensity of the current and R is the resistance of the circuit. This gives, as the total change of induction, the quantity

$$R\int i dt,$$

if we assume that R is constant. But this is equal to Rq, where q is the total quantity of electricity which has passed. Its value may, therefore, be determined experimentally by means of the indications of the galvanometer. (It must be remembered that if the coil consist, for example, of n turns wound closely on the bar, the actual induction—supposed to be uniform—in the bar is Rq/n.) Hence (§§ 351, 352) we can determine the permeability and susceptibility of the substance of which the bar is composed.

370. *Electric and Magnetic Units.*—The magnitude of electric and magnetic, as of all other quantities, depends upon the particular units in terms of which they are measured. All such quantities may be expressed in terms of the units of mass, length, and time; but the *dimensions* of a quantity in terms of these units depends upon the particular definition of some electric or magnetic quantity which we adopt.

Two systems of measurement are in use—the *Electrostatic* and the *Electromagnetic*. In the electrostatic system we start from the definition that two similar unit quantities of electricity, condensed at points which are at unit distance apart, repel each other (in air) with unit force (§ 312).

The dimensions of force are (§ 64) (MLT^{-2}), and, therefore, the dimensions of electric quantity are

$$(q) = (M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}).$$

Surface density of electricity is quantity per unit surface. Its dimensions are, therefore, on this system,

$$(\sigma) = (M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1}).$$

Electric potential and electric force have dimensions (§ 313)

$$(v)=(qL^{-1})=(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$$

and

$$(qL^{-2})=(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$$

respectively.

The dimensions of electrostatic capacity are (§ 314)

$$(qv^{-1})=(L):$$

those of current-strength are (§ 335)

$$(qT^{-1})=(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}):$$

and those of resistance are (§ 336) directly proportional to those of potential and inversely proportional to those of current-strength; they are, therefore,

$$(R)=(L^{-1}T).$$

On the electromagnetic system the definition of unit quantity of magnetism is precisely analogous to the definition of unit quantity of electricity in the electrostatic system. Hence the dimensions of quantity of magnetism, surface density of magnetism, magnetic potential, and magnetic force, are, on this system, identical with the dimensions of the corresponding electric quantities on the electromagnetic system.

In addition, the dimensions of magnetic moment and intensity of magnetisation on the latter system are (§§ 349, 351)

$$(m)=(qL)=(M^{\frac{1}{2}}L^{\frac{5}{2}}T^{-1})$$

and

$$(\mathfrak{I})=(qL^{-2})=(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$$

respectively. The latter expression, of course, is identical with the expression for the dimensions of surface density.

On the electromagnetic system, unit current is the current which, flowing in a circular circuit of unit radius, exerts unit force per unit length of its circumference upon a unit magnetic pole placed at its centre (§ 365). Hence the dimensions of current-strength are

$$(C)=(MLT^{-2}q^{-1}L)=(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1}).$$

The quantity of electricity which is conveyed through a conductor is directly proportional to the strength of the current and to the time during which it has been flowing. Therefore, its dimensions are

$$(Q)=(CT)=(M^{\frac{1}{2}}L^{\frac{1}{2}}).$$

The dimensions of electric potential when multiplied by quantity of electricity are (§ 320) identical with those of energy, and are, therefore,

$$(V) = (ML^2T^{-2}Q^{-1}) = (M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2}).$$

By such considerations we may readily determine the dimensions of any electrical or magnetic quantity on either system of reckoning. Some of the results are tabulated below, the dimensions on the electrostatic system being given in the second column, while those on the electromagnetic system are given in the third.

Electrical Quantities.

Quantity of Electricity.....	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}})$
Surface density of Electricity	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{-\frac{3}{2}})$
Electric Potential	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2})$
Electric Force	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-2})$
Electrostatic Capacity	(L)	$(L^{-1}T^{2})$
Current Strength	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-2})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$
Resistance	$(L^{-1}T)$	(LT^{-1})
Specific Inductive Capacity...	$(M^{0}L^{0}T^{0})$	$(L^{-2}T^{2})$

Magnetic Quantities.

Quantity of Magnetism	$(M^{\frac{1}{2}}L^{\frac{1}{2}})$	$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1})$
Surface density of Magnetism	$(M^{\frac{1}{2}}L^{-\frac{3}{2}})$	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$
Magnetic Potential	$(M^{\frac{3}{2}}L^{\frac{3}{2}}T^{-2})$	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-1})$
Magnetic Force.....	$(M^{\frac{1}{2}}L^{\frac{1}{2}}T^{-2})$	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$
Magnetic Moment	$(M^{\frac{1}{2}}L^{\frac{3}{2}})$	$(M^{\frac{1}{2}}L^{\frac{5}{2}}T^{-1})$
Intensity of Magnetisation ...	$(M^{\frac{1}{2}}L^{-\frac{3}{2}})$	$(M^{\frac{1}{2}}L^{-\frac{1}{2}}T^{-1})$
Magnetic Permeability	$(L^{-2}T^{2})$	$(M^{0}L^{0}T^{0})$
Magnetic Susceptibility	$(L^{-2}T^{2})$	$(M^{0}L^{0}T^{0})$

It is specially worthy of notice that the dimensions of any quantity on one or other of these systems always differ from its dimensions on the other by the dimensions of a speed or of a speed squared.

The force between two quantities, q and q', of electricity at a distance r apart in a medium other than air, is $qq'Kr^2$, where K is the specific inductive capacity. Hence, if we choose not to define the electrostatic dimensions of K as zero, but leave them undetermined, the electrostatic dimensions of quantity of electricity become

$$(M^{\frac{1}{2}}L^{\frac{3}{2}}T^{-1}K^{\frac{1}{2}}).$$

Similarly, if we leave the dimensions of magnetic permeability (μ) undetermined, we find that the electromagnetic dimensions of quantity of electricity are

$$(M^{\frac{1}{2}}L^{\frac{1}{2}}\mu^{-\frac{1}{2}});$$

and the corresponding alterations in the dimensions of other quantities can easily be found. One advantage of this method (due to Rücker) is, as Fitzgerald pointed out, that we can make the dimensions of any one quantity on both systems identical by assuming that the dimensions of K and μ are (TL^{-1}).

It is convenient, in scientific measurements, to adopt the centimetre-gramme-second (c.g.s.) system of units; but, in practice, these units are often inconveniently large or inconveniently small. In the following table the name of the practical unit of various quantities is given in the second column; and the factors which are required to reduce the numerics, as expressed in terms of the practical units, to their equivalents on the c.g.s. system, are given in the third column.

Quantity of Electricity	Coulomb	10^{-1}
Electromotive Force	Volt	10^{8}
Electrostatic Capacity	Farad	10^{-9}
	Microfarad	10^{-15}
Current Strength	Ampère	10^{-1}
Resistance	Ohm	10^{9}

An electromotive force of one volt maintains a current, whose strength is one ampère, through a resistance of one ohm.

CHAPTER XXXII.

ELECTROMAGNETIC THEORY OF LIGHT.

371. *Magnetic Rotation of the Plane of Polarisation of Light.*—Faraday made many attempts to detect some action upon polarised light when it was made to pass through a dielectric which was subjected to electric stress. He also sought for evidence of such action when polarised light passed through an electrolyte conveying a current, but in no case could he observe any effect. On the other hand, he found a marked effect when polarised light was passed through a diamagnetic medium placed in a field of magnetic force.

When the direction of the ray coincides with the positive direction of the lines of force, the plane of polarisation is rotated through an angle which is proportional to the intensity of the magnetic field, and to the length of the path of the ray within the medium. If the direction of the ray does not coincide with the direction of the field, the rotation is proportional to the intensity of the resolved part of the force taken in the direction of the ray. The amount of the rotation per unit of length, in a field of unit intensity, depends upon the nature of the medium. The absolute direction of rotation is unaltered by a reversal of the ray, provided that the direction of the field is unaltered.

In diamagnetic media the direction of the rotation is, in general, connected with that of the field according to the law of right-handed screwing motion; in paramagnetic media, the reverse is generally true.

The fact of the non-reversal of the rotation with reference to the direction of the field points to a fundamental distinction between the mechanical method by which this magnetic rotation is produced and that obtained in cases of rotation by quartz or solutions such as sugar (§ 251). In the latter cases, reversal of the ray is not accompanied by a reversal of the rotation with reference to *it;* and thus the total rotation during the double passage through the medium

is zero. The total magnetic rotation is doubled by the double passage.

In the case of vapours and gases, the rotation is very small in comparison with the rotation which is produced, under similar conditions, by solids and liquids. Even in vapours of liquids which have considerable rotatory power, the effect is very small.

Verdet has shown that the rotation is approximately inversely proportional to the square of the wave-length—the deviation being in *defect* as the wave-length increases, and being most marked in substances of great dispersive power.

372. *Hypothesis of Molecular Vortices.*—We have seen that a plane polarised ray may be compounded of two uniform equiperiodic circular motions of equal amplitude, and that rotation of the plane of polarisation will take place if one of these component motions is accelerated relatively to the other (§ 251). The explanation of the magnetic effect seems to lie in this direction.

Sir W. Thomson has remarked on this subject 'That the magnetic influence on light discovered by Faraday depends on the direction of motion of moving particles. For instance, in a medium possessing it, particles in a straight line parallel to the lines of force, displaced to a helix round this line as axis, and then projected tangentially with such velocities as to describe circles, will have different motions according as their motions are round in one direction (the same as the nominal direction of the galvanic current in the magnetising coil), or in the contrary direction. But the elastic reaction of the medium must be the same for the same displacements, whatever be the velocities and directions of the particles; that is to say, the forces which are balanced by centrifugal force of the circular motions are equal, while the luminiferous motions are unequal. The absolute circular motions being therefore either equal or such as to transmit equal centrifugal forces to the particles initially considered, it follows that the luminiferous motions are only components of the whole motion, and that a less luminiferous component in one direction, compounded with a motion existing in the medium when transmitting no light, gives an equal resultant to that of a greater luminiferous motion in the contrary direction compounded with the same non-luminous motion. I think it not only impossible to conceive any other than this dynamical explanation of the fact that circularly-polarised light transmitted through magnetised glass parallel to the lines of magnetising force, with the same quality, right-handed always, or left-handed always, is propagated at different rates according as its course is in the direction, or is contrary to the direction, in which a north magnetic pole is

drawn; but I believe it can be demonstrated that no other explanation of that fact is possible. Hence it appears that Faraday's optical discovery affords a demonstration of the reality of Ampère's explanation of the ultimate nature of magnetism; and gives a definition of magnetisation in the dynamical theory of heat. The introduction of the principle of moments of momenta ("the conservation of areas") into the mechanical treatment of Mr. Rankine's hypothesis of "molecular vortices" (see § 254), appears to indicate a line perpendicular to the plane of resultant rotatory momentum ("the invariable plane") of the thermal motions as the magnetic axis of a magnetised body, and suggests the resultant moment of momenta of these motions as the definite measure of the "magnetic moment." The explanation of all phenomena of electromagnetic attraction and repulsion, and of electromagnetic induction, is to be looked for simply in the inertia and pressure of the matter of which the motions constitute heat. Whether this matter is or is not electricity, whether it is a continuous fluid interpermeating the spaces between molecular nuclei, or is itself molecularly grouped; or whether all matter is continuous, and molecular heterogeneousness consists in finite vortical or other relative motions of contiguous parts of a body, it is impossible to decide, and perhaps in vain to speculate, in the present state of science.'

The idea contained in these remarks has been developed by Maxwell into a complete theory of molecular vortices. He points out that, from the fact that the wave-length, λ, and the periodic time, τ, increase and decrease together, it follows that if for a given numerical value of the angular velocity, n $(=2\pi/\tau)$, the value of the speed of propagation, λ/τ, is greater when n is positive than when it is negative, for a given value of λ the positive value of n will be greater than the negative value. This is so since the former condition implies that λ is greater when n is positive than when it is negative, and a diminution of λ implies a diminution of τ and therefore an increase of n. Since the ray does not diminish in intensity as it passes through the medium, the amplitude, r, must remain constant (§ 179); and the principle of conservation of energy shows that, for equilibrium, we must have the condition

$$-\frac{d\mathrm{T}}{dr}+\frac{d\mathrm{V}}{dr}=0,$$

where T and V are respectively the kinetic and potential energies. But the expression for T contains one term involving n^2; and it may contain terms involving the products of n into other velocities, and

terms independent of n. V, on the other hand, is independent of n. Hence the above equation is of the form

$$An^2 + Bn + C = o,$$

where A, B, and C, are functions of the co-ordinates. Now experiment shows that n has two real values, one positive, the other negative and smaller. C must therefore be finite, and both it and B must be negative if A is positive; for $-B/A$ and C/A are respectively the sum and the product of the roots of the equation, and the sum is positive while the product is negative. B also cannot vanish, since the roots are distinct. The term in n must therefore involve another velocity besides n, and that velocity must be an angular velocity about the same axis, for Bn is a scalar quantity.

Maxwell then concludes 'That in the medium, when under the action of magnetic force, some rotatory motion is going on, the axis of rotation being in the direction of the magnetic forces; and that the rate of propagation of circularly polarised light, when the direction of its vibratory rotation and the direction of the magnetic rotation of the medium are the same, is different from the rate of propagation when these directions are opposite.

'The only resemblance which we can trace between a medium through which circularly-polarised light is propagated, and a medium through which lines of magnetic force pass, is that in both there is a motion of rotation about an axis. But here the resemblance stops, for the rotation in the optical phenomenon is that of the vector which represents the disturbance. This vector is always perpendicular to the direction of the ray, and rotates about it a known number of times in a second. In the magnetic phenomenon, that which rotates has no properties by which its sides can be distinguished, so that we cannot determine how many times it rotates in a second.

'There is nothing, therefore, in the magnetic phenomenon which corresponds to the wave-length and the wave-propagation in the optical phenomenon. A medium in which a constant magnetic force is acting is not, in consequence of that force, filled with waves travelling in one direction, as when light is propagated through it. The only resemblance between the optical and the magnetic phenomenon is, that at each point of the medium something exists of the nature of an angular velocity about an axis in the direction of the magnetic force.'

'This angular velocity cannot be that of any portion of the medium of sensible dimensions rotating as a whole. We must therefore con-

ceive the rotation to be that of very small portions of the medium, each rotating on its own axis. This is the hypothesis of molecular vortices.

'The motion of these vortices, though . . . it does not sensibly affect the visible motions of large bodies, may be such as to affect that vibratory motion on which the propagation of light, according to the undulatory theory, depends. The displacements of the medium, during the propagation of light, will produce a disturbance of the vortices, and the vortices when so disturbed may react on the medium so as to affect the mode of propagation of the ray.'

From this hypothesis Maxwell has deduced an expression for the magnitude of the rotation under given conditions which accords very well with the results of observation.

373. *Hall's Effect.*—Hall found by experiment that a thin metallic conductor, which is placed in a magnetic field of force with its plane perpendicular to the lines of force, and through which an electric current flows, is the seat of an electromotive force which acts along the common perpendicular to the directions of the current and the magnetic field. Rowland has proved that, if a similar electromotive force appears when the electric displacement in an insulating medium varies in a field of magnetic force, rotation of the direction of the displacement will follow the passage of a wave through the medium; and Glazebrook has shown that the electromotive force is a consequence of the molecular rotation which Maxwell assumes.

374. *Kerr's Effects.*—The action upon polarised light in a medium subjected to electric stress, for which Faraday sought in vain, was discovered by Kerr. He placed two parallel brass plates at a short distance apart, in a glass cell containing carbon bisulphide, and connected them with the poles of an electric machine. A beam of light, polarised at an angle of 45° to the direction of the lines of electric force, was passed between the plates, and, on emergence, was found to be elliptically polarised. The axes were respectively parallel to and perpendicular to the lines of force; and the difference between the phases of the two components was proportional to the square of the intensity of the electric field.

Another form of this experiment consists in passing the polarised beam between two small spheres which are placed near to each other in the insulating medium, and are connected to the poles of an electric machine. The medium is then found to have become doubly refractive.

Dr. Kerr has also discovered that the plane of polarisation of light is rotated in the act of reflection from the (highly-polished) pole of

an electromagnet—the direction of the rotation being reversed when the magnetisation of the pole is reversed.

375. *Electromagnetic Theory of Light.*—The phenomena which are described in the immediately preceding sections point to a very close connection between electricity, magnetism, and light.

We have already seen that the phenomena of light are best explained on the assumption that light consists in undulations propagated through a medium. On the other hand, the theories of electrical and magnetic action were originally expressed in terms of direct action at a distance; and (although Faraday conducted all his reasoning on the assumption of action through a medium) it was not until Maxwell translated Faraday's ideas into mathematical language that it was recognised that electrical and magnetic phenomena could be readily explained as the results of the propagation of action through a medium.

In determining the conditions of the propagation of an electromagnetic disturbance through the medium whose existence he postulated, Maxwell arrived at the conclusion that the propagation takes place in accordance with the laws of the transference of motion through an elastic solid, and that the speed of propagation is

$$V = \frac{1}{\sqrt{K\mu}},$$

where K and μ are respectively the specific inductive capacity and the permeability of the medium. In the propagation of a plane wave, electric displacement takes place at right angles to the direction of magnetic induction, and both are in the plane of the wave.

A reference to § 370 will make it evident that, on either of the electrostatic or the electromagnetic systems, the dimensions of K are those of the inverse square of a speed. According to Maxwell, if light is an electromagnetic phenomenon, V must represent the speed of light. Now the speed of light, v, is capable of measurement to a considerable degree of accuracy, while the value of V can be determined directly by a comparison of the relative values of some electrical or magnetic quantity on the two systems of measurement (§ 370); and the results of various independent determinations of the values of V and v strongly confirm the supposition of their numerical identity in air.

In media other than air the speed of light is inversely proportional to the refractive indices. Hence K should be practically equal to the square of the refractive index, for the value of μ is nearly unity in all transparent media. Since the experimental

determination of K occupies a time which is practically infinite in comparison with the period of any luminous vibration, we must, in testing this point, take the refractive index for rays of infinite wave-length. (This might be given by the value of the constant a in Cauchy's expression for the refractive index, § 209). Hopkinson has found that while the relation holds in the case of hydrocarbons, it does not obtain in glass and the animal and vegetable oils. In the latter substances the refractive index is less than $\sqrt{K}$.

In the luminiferous medium, two forms of energy exist—one kinetic, the other potential. Similarly, in the electromagnetic medium, energy exists in a kinetic (electrokinetic, § 368) form and in a potential (electrostatic, § 320) form.

The theory also explains double refraction, and leads to Fresnel's construction for the wave surface. And it shows that the speed of propagation of a condensational-rarefactional wave would be infinite; so that this wave does not exist—a result which is in harmony with optical observations. Also, if the medium be not a perfect conductor, the electrical energy is in part transformed into energy of electric currents, and so finally into heat. This explains the absorption of light.

376. *Electromagnetic Waves.*—The above evidence in favour of the truth of Maxwell's electromagnetic theory is very strong in itself. But recent investigations, by Hertz and others, have proved, beyond the possibility of doubt, that electromagnetic action is propagated with finite speed through a medium, and have indicated that its speed of propagation is identical with that of light.

If the initial disturbance is periodic, a series of electromagnetic waves are propagated outwards from the source. The condition of periodicity can be obtained by means of the disruptive discharge, which, under suitable conditions, is oscillatory in its nature (§ 321) and has a constant period of oscillation depending on the electrostatic capacity and the coefficient of self-induction of the apparatus which is used for the production of the discharge.

Let us suppose, for the sake of definiteness, that the discharge takes place between the poles of a Holtz machine. The alternating currents which characterise the discharge will induce similar currents in neighbouring conductors. According to the theory of direct action at a distance, these induced currents will appear in exact simultaneity with the inducing currents; according to the electromagnetic theory, they will appear later and later in proportion as the conductors in which they are induced are more and more remote from the Holtz machine.

In his investigations on this point, Hertz made use of the

principle of resonance (*cf.* § 173). That is, he used a secondary conducting circuit the natural period of electric oscillation in which was the same as that in the primary circuit. The result is that the magnitude of the induced oscillations may be made very great, for each succeeding induced oscillation is timed to re-enforce the effects of preceding oscillations. In this way sufficient electromotive force may be developed to cause the electricity to spark across a small air-gap in the circuit.

Such a circuit, with its air-gap, may be used to make evident the existence of inductive effects at any given point in space, provided that its distance from the source is not too great. If one source alone existed in space, the intensity of the inductive effect at any given point, and therefore the intensity of the spark in the secondary circuit, would diminish continuously as the distance between the point and the source increased. If two sources existed, the intensity might be great in the neighbourhood of each and might reach a minimum at some intermediate position: or, if the effects of the sources were opposite, the intensity of the resultant effect might be zero at some point between the two. If the effects were instantaneously propagated, only one such minimum could exist. But, if the effects were propagated by wave-motion at a finite rate, a great number of maxima and minima might appear, in accordance with the ordinary laws of interference. The best results will be obtained when the two sources are precisely similar. Now, if a conducting sheet be placed in the neighbourhood of a single source, the currents which are induced in it will, in turn, give rise to electromagnetic effects having a periodic time equal to that of the source. This, on the electromagnetic theory, constitutes reflection of the electromagnetic radiation, and interference may be expected to take place between the incident and the reflected waves —nodes and loops occurring alternately at equal intervals of one half of the length of a wave (§ 53). In performing this experiment Hertz was able to observe the existence of successive maxima and minima, and so the existence of electromagnetic radiation was proved.

If the original electrical oscillations take place along a straight rod, the oscillations in the electromagnetic medium will be parallel to the axis of the rod; *i.e.*, the wave is plane polarised. And the rod is surrounded by circular lines of magnetic induction, the direction of the induction changing with each alternation of electric displacement. The electric displacement and the magnetic induction therefore take place in the front of the wave, and are at right angles to each other; and these two effects can be separated from

each other by using a suitable resonating circuit. A circular circuit held with its plane passing through the axis of the rod and its spark-gap at right angles to the axis will respond only to the magnetic variations. On the other hand, if held in front of the axis with the line of its spark-gap and one of its diameters parallel to the axis, it will respond only to the electrical variations, for no lines of magnetic induction pass through it.

With apparatus such as is ordinarily used in a laboratory, radiations having a wave-length varying from a few inches to a number of miles in length, can be readily obtained. These long-period radiations can pass freely through insulators, such as pitch, which are opaque to luminous radiations. And, by using large prisms of such substances, their refractive indices can be found by the usual methods.

If the radiation falls upon a transparent sheet of thickness which is small in comparison with the wave-length, no reflection is observed; for the acceleration of the phase by half-a-period, which takes place at the second surface, produces total interference. This is a reproduction of the phenomenon of the central black spot in Newton's rings (§ 221).

Reflection can be obtained at the surface of a thick insulator if the direction of the electric displacements is perpendicular to the plane of reflection; but none is found at the polarising angle if the line of displacement lies in the plane of reflection. This proves that the electric displacement takes place at right angles to the plane of polarisation, and settles the much debated question of the direction of the luminous vibrations in favour of Fresnel's assumption (§ 239).

Effects of diffraction can also be observed with these waves, and are in strict accordance with the results of the undulatory theory.

Maxwell concluded from his theory that a body which absorbs light should be repelled towards the unilluminated side. The effect is too small for observation with luminous rays—even with concentrated sunlight; but a disc of good conducting silver is repelled from the pole of an electromagnet which is excited by a powerful alternating current.

CHAPTER XXXIII.

THE ETHER.

377. Whenever mutual action is observed between two systems, it is possible to explain the accompanying phenomena, in great part at least, on either of two assumptions. We may assume that the action occurs directly at a distance, or we may assume that it is propagated by means of a material medium. But when we can show that the action takes time to travel from one point of space to another, the latter assumption only is tenable. For example, we have no evidence as yet (§ 92) that gravitational action is not instantaneously propagated, and therefore either assumption is valid; but the experiments of Hertz have shown that electrodynamic action requires a finite time for its propagation over a finite space, and therefore we must assume the existence of a medium by means of which that action is transferred.

Such a medium is termed an 'ether,' and we may therefore define the ether as a substance, other than ordinary matter, through which action is propagated.

At one time many ethers were supposed to exist—indeed, a new ether was postulated for the explanation of almost every new class of phenomena which presented itself. An ether was assumed in order to explain gravitation. Newton introduced a medium to account for the production of the 'fits' of easy reflection and transmission of luminous corpuscles, which he had to postulate for the explanation of the colours of thin plates, etc., on the corpuscular theory. The aid of another was invoked for the explanation of physiological phenomena, and so on—the attributes with which each was endowed being specially chosen in order to make it suit each particular case.

Such a procedure is totally unscientific, and it is now the aim of scientists to ascribe all actions, which apparently occur directly at a distance, to the intervention of a single medium. Of all the host of mediæval ethers, one alone remains—the medium whose existence was postulated by Huyghens in his explanation of the phenomena

of light; and one of the great merits of Maxwell's modern electromagnetic ether is that it explains the phenomena of light as well as the phenomena of electricity and magnetism.

378. It is more easy to say what the ether is *not* than to say what it is. It is not a gas, like air; for transverse oscillations, such as those which take place in the propagation of light, die out with extreme rapidity in such a medium. Neither, for a like reason, is it a liquid like water. So far as this effect is concerned, it might be a transparent solid, for solids can transmit transverse oscillations. But on the other hand, the rate at which transparent solids transmit such vibrations is immensely slower than the rate at which the ether transmits light. Hence the ether cannot be an ordinary transparent solid, although it interpenetrates such solids, and is hampered in its action by them—a fact shown by the diminished speed of light when passing through them.

Yet the ether, although it is not ordinary matter, must be material, *i.e.*, must possess inertia, for it transmits energy at a finite rate. And, for the same reason, it must possess rigidity and must be elastic. The rate of vibration of the parts of the medium is (*cf.* § 168) directly proportional to the square root of the rigidity, and is inversely proportional to the square root of the density. When red light passes through the ether, the rate of vibration is about 400,000,000 times per second. A steel tuning-fork which emits even the highest audible note would require to be immensely more rigid than it is if it were to vibrate at that rate; if its rigidity remained constant, it would require to be far less massive than it is. It would appear from Thomson's calculations, based on a very plausible assumption, that the ether is about $(10)^9$ times less rigid than steel is; but, on the other hand, its density would appear to be about $(10)^{19}$ times less than that of steel.

379. It seems, therefore, that the ether acts as if it were an elastic solid. And yet the earth, in its course round the sun, moves through it without being subjected to any appreciable resistance; and the light which comes to us from a distant star gives no evidence of the disturbing effect which the earth might be supposed to have upon the ether in its neighbourhood as it moves quickly through it.

Young therefore suggested that the structure of the earth and other solid matter is such that the ether flows freely through it, being subjected to no alteration other than a change of density. But Stokes has shown that this rather startling assumption is probably unnecessary, if the speeds of the earth and of the particles of air in its atmosphere are small in comparison with that of light.

Of course, the free motion of the earth through the ether shows that, relatively to the moving earth, the ether acts like a practically non-viscous fluid, and we have to explain how it is that the ether can act both like a fluid and like an elastic solid. No explanation of this point can be more lucid than the original explanation given by Stokes: 'The plasticity of lead is greater than that of iron or copper, and, as appears from experiment, its elasticity is less. On the whole it is probable that the greater the plasticity of a substance, the less its elasticity, and *vice versâ*, although this rule is probably far from being without exception. When the plasticity of the substance is still further increased, and its elasticity diminished, it passes into a viscous fluid. There seems no line of demarcation between a solid and a viscous fluid. In fact, the practical distinction between these two classes of bodies seems to depend on the intensity of the *extraneous* force of gravity, compared with the intensity of the forces by which the parts of the substance are held together. Thus, what on the Earth is a soft solid might, if carried to the Sun, and retained at the same temperature, be a viscous fluid, the force of gravity at the surface of the Sun being sufficient to make the substance spread out and become level at the top; while what on the Earth is a viscous fluid might on the surface of Pallas be a soft solid. The gradation of viscous into what are called perfect fluids seems to present as little abruptness as that of solids into viscous fluids; and some experiments which have been made on the sudden conversion of water and ether into vapour, when enclosed in strong vessels and exposed to high temperatures, go towards breaking down the distinction between liquids and gases.

'According to the law of continuity, then, we should expect the property of elasticity to run through the whole series, only it may become insensible, or else may be mastered, by some other more conspicuous property. It must be remembered that the elasticity here spoken of is that which consists in the tangential force called into action by a displacement of continuous sliding; the displacements also which will be spoken of in this paragraph must be understood to be such displacements as are independent of condensation or rarefaction. Now, the distinguishing property of fluids is the extreme mobility of their parts. According to the views explained in this article, this mobility is merely an extremely great plasticity, so that a fluid admits of a *finite*, but exceedingly small amount of constraint before it will be relieved from its state of tension by its molecules assuming new positions of equilibrium. Consequently the same oblique pressures can be called into action in a fluid as in a solid, provided the amount of relative displacement of the parts be

exceedingly small. All we know for certain is that the effect of elasticity in fluids [elasticity of form] is quite insensible in cases of equilibrium, and it is probably insensible in all ordinary cases of fluid motion. . . . But a little consideration will show that the property of elasticity may be quite insensible in ordinary cases of fluid motion, and may yet be that on which the phenomena of light entirely depend. When we find a vibrating string, the small extent of vibration of which can be actually seen, filling a whole room with sound, and remember how rapidly the intensity of the vibrations of the air must diminish as the distance from the string increases, we may easily conceive how small in general must be the amount of the relative motion of adjacent particles of air in the case of sound. Now, the extent of the vibration of the ether in the case of light may be as small, compared with the length of a wave of light, as that of the air is compared with the length of a wave of sound; we have no reason to suppose it otherwise. When we remember, then, that the length of a wave of sound in air varies from some inches to several feet, while the greatest length of a wave of light is about ·00003 of an inch, it is easy to imagine that the *relative* displacement of the particles of ether may be so small as not to reach, nor even come near to, the greatest relative displacement which could exist without the molecules of the medium assuming a new position of equilibrium, or, to keep clear of the idea of molecules, without the medium assuming a new arrangement which might be permanent.'

In this connection Thomson refers to the properties of shoemakers' wax, which is so brittle that it will splinter under a sudden blow, and which will flow like a liquid into all the crevices of the vessel which contains it, while leaden bullets will sink down through it, and corks will float up through it—if only sufficient time be allowed (§ 78). The resistance to the passage of a bullet or a cork through it becomes smaller and smaller, the slower the motion becomes; and it may be that the motion of the earth through the ether is far less, relatively to the resisting power of the ether, than is the motion of the bullet or the cork relatively to the resisting power of the wax.

380. The above theory of the ether is known as the elastic-solid theory. This elastic solid cannot possess positive compressibility, for in that case a condensational-rarefactional wave — of whose existence we have no experimental evidence — might be propagated through it with finite speed. Hence Green, who investigated the properties of this ether, assumed that it was incompressible. He recognised the case of negative compressibility, but dismissed it with

the assertion that a medium which possesses negative compressibility—*i.e.*, a medium which expands when subjected to increased pressure, and contracts when pressure is removed—is necessarily unstable.

In order, on this theory, to account for the reflection and refraction of light at an interface, Green assumed that the ether had the same rigidity, but was of unequal density, on the two sides of the interface. This gave Fresnel's law in the case of vibrations perpendicular to the plane of reflection; but, in the case of vibrations in the plane of reflection, it gave a result which only coincided with Fresnel's law when the refractive indices of the two media were practically identical.

Sir W. Thomson assumed that the ether consists of an inviscid fluid permeating the pores of an incompressible sponge-like solid: but the result deviated further from Fresnel's law than Green's did. He therefore had to abandon the doctrine of incompressibility; and, having pointed out that Green's negatively compressible medium was not unstable if it were infinite or had rigid boundaries, he assumed such negative compressibility as to make the velocity of the condensational-rarefactional wave zero. He assumed (like Green) equal rigidities of the ether in the media; but this condition has been shown to be necessary for stability when the other assumed conditions hold.

This contractile ether gives Fresnel's laws; and Glazebrook has shown that it explains the reflection and refraction of light by transparent bodies and by metals, double refraction and dispersion (including anomalous dispersion), and that it gives the correct expression for the velocity of light in a moving medium.

Thomson has also shown how a model of a medium might be constructed by rigid jointed connections and rigid revolving flywheels (or gyrostats in which a frictionless fluid circulated irrotationally) which has no intrinsic rigidity, *i.e.*, no intrinsic elastic resistance to change of shape, but which has a quasi-rigidity due to inherent resistance to rotation; which is absolutely devoid of resistance to change of volume or to irrotational change of shape; which therefore is incapable of transmitting condensational-rarefactional waves, but which can transmit vibrations like those of light. It is therefore a practical realisation of his contractile ether.

381. The electromagnetic theory of the ether has been discussed in the last chapter. It readily explains all the difficulties which originally beset the elastic-solid theory; and the question of the condensational-rarefactional wave never arises, for its velocity of propagation is infinite. On some points its results differ from those

of Thomson's theory, but the differences are too small to admit of crucial tests being based upon them.

It must be observed, however, that this theory stands upon a different footing from the former. No fundamental assumption is made regarding the medium other than that it shall account for certain electrical and magnetic actions.

382. The property of *dilatancy* in a medium composed of rigid particles in contact accounts for a number of natural phenomena and presents analogies to many others.

Let us suppose that we have a space filled with marbles or shot, each being in contact with another on various sides. This condition can be satisfied by different arrangements of the spheres, so that in some arrangements the volume occupied by a given number is less than the volume occupied by the same number in other arrangements. There is an arrangement of maximum volume and an arrangement of minimum volume, and we cannot change the hard spheres from one arrangement to another without altering the volume. (This explains the meaning of the term 'dilatancy.') Change of shape of such a mass of spheres cannot occur without simultaneous change of bulk. Hence, if the mass be enclosed in an inextensible, but flexible, boundary, no change of shape which necessitates change of volume can occur.

If the mass of spheres is enclosed by a smooth boundary, motion of the layer next the boundary will cause less alteration of volume than does the motion of a layer in the interior of the mass. Hence, when certain stresses are applied, there may be a streaming motion of the spheres along the boundary, while the rest of the spheres do not move. This conduction of the parts of the medium along a smooth surface resembles the conduction of electricity.

If, in a large mass of spheres in the condition of maximum density enclosed in an elastic boundary, one sphere grows in size, the whole medium at first undergoes dilatation. Then the layer next the growing sphere reaches the condition of maximum volume. After that, the layer next the sphere will be returning to the condition of minimum volume, while a layer a little farther out is at maximum volume. Later, there will be a succession of maxima and minima in the neighbourhood of the growing body.

When two bodies, growing in size, are present in the medium at a considerable distance apart, the resultant dilatation, at any point, due to both is less than the sum of the separate dilatations at that point due to each. Thus there will be a force of attraction between the bodies whose magnitude depends upon the rate at which the dilatation varies with the distance between the bodies. The dilata-

tion becomes periodic when the bodies are near to each other, and attraction and repulsion occur alternately. This resembles the phenomena of molecular forces.

Instead of supposing that the boundary is elastic, we may assume that it is rigid, and that the growing spheres are elastic. We may even suppose that the spheres are rigid provided that the medium is composed of large spheres scattered uniformly among small spheres, for such a medium may possess elasticity in virtue of the propagation of distortional waves through it—just as a slack chain possesses elasticity when lateral vibrations are passing along it. Even if the small grains were at maximum density, distortional waves could pass, the distortions of the two sets of grains being opposite. This may throw light on electrodynamic and magnetic phenomena. Also, the separation of the two sets of grains would produce phenomena analogous to those depending on the separation of positive and negative electricities. And, with a certain arrangement of the large and small grains, the state of stress in the medium is the same as that which must exist in the ether in order to account for gravity.

Such, in brief outline, is Osborne Reynolds' theory of a granular ether.

383. It has been shown also that an ether consisting of vortices in a perfect fluid might be capable of transmitting light. And the instantaneous propagation of gravitational action (if it be instantaneous) does not in this case present so great difficulty; for, in a certain sense, each vortex occupies all the space which the fluid fills—its action is instantaneously felt in all parts.

384. It is not to be supposed that all these theories of the ether are necessarily antagonistic. The vortex theory, for example, may be the same as the elastic-solid theory.

INDEX.

THE NUMBERS REFER TO THE SECTIONS.

Aberration, chromatic, 196
spherical, 196
Absorption, co-efficient of, 205
Absorptive power, 203, 256
Acceleration, 42, 43, 44, 46, 50
Accumulator, water-dropping, 322
Achromatism, 197
Actinometer, 260
Activity, 68
Adiabatics, 295
Amplitude, 51
Analogy, 15
Atmolysis, 112
Atom, 82, 139-141
Avogadro's Law, 150

Beats, musical, 175
Biprism, 214
Boiling, 275
Bolometer, 343
Boyle's Law, 103, 105, 149, 150
Brewster, 203, 208, 217, 223, 238, 243, 251, 263
Brittleness, 83

Caloric, 254
Capillarity, 119-127
Carnot's cycle, 290, 291
Cauchy, 146, 209, 258
Caustics, 184, 189
Cavendish experiment, 89
Characteristic function, 201
Charles' Law, 150, 265
Chemical combination, 280
Cohesion, 118, 137
Colour, 180
body, 205
surface, 206
Colours of crystalline plates, 248, 249
mixed plates, 222
thick plates, 223
thin plates, 218-221
Compressibility, 79
of gases, 102
of a perfect gas, 104
of vapours, 106
of liquids, 113
of solids, 128
Conductivity, thermal, 151
Conservation, 4, 12
of energy, 4, 10
of matter, 4, 5
Consonance, 175
Contact, angle of, 121
Coronæ, 232
Coulomb's torsion balance, 312
Couple, 70
Critical angle of reflection, 188
Crucial test, 15
Crystalline structure, 143, 144
Curvature, 43

Density, 64
of earth, 89, 90
Deviation, minimum, 192
Dew, 277
Dialysis, 117
Diamagnetism, 346
Dichroism, 205
Dielectrics, 306
Diffraction gratings, 233
Diffusion of gases, 109, 151
of liquids, 116
Diffusivity, 109
Dilatancy, 382

Dilatation of gases, 265
of liquids, 264
of solids, 262
Dimensions, 18, 67
Discharging rod, 314
Dispersion, irrationality of, 197
theories of, 209
Displacement, 40
Disruptive discharge, 321
Dissociation, 152, 280
Dissonance, 175, 176
Divisibility, 79, 139
Döppler's principle, 178
Dulong and Petit, law of, 270
Dynamical similarity, 67, 73, 76, 127, 168, 170, 171
Dynamics, 63
Dynamo, 367

Efficiency, thermo-dynamic, 293
Effusion, 110
Elasticity, 79
of gases, 107
of liquids, 114
of solids, 134, 156
Electric absorption, 321
charge, 310
current, 335, 343
density, 316
displacement, 319
force, lines of, 318
images, 317
induction, 310, 319
machines, 326
polarisation, 338
potential, 313
quantity, 310
resistance, 335, 343
Electrodynamometer, 369
Electrokinetic energy, 368
Electrolysis, 337
Electromagnetic induction, 367
medium, 92
units, 370
waves, 376
Electrometer, 325
Electromotive force, 313, 343
Electrophorus, 311
Electroscope, gold-leaf, 309
Electrostatic capacity, 314
energy, 320
units, 370
Emissivity, 203, 256
Empirical formulæ, 17
laws, 17
Energy, 4, 7-11, 62
conservation of, 10
dissipation of, 11, 297, 298
forms of, 8
kinetic, 8
potential, 8, 11
transformation of, 9, 11
Entropy, 296, 298
Epoch, 51
Equation, personal, 16
Equilibrium, of fluids, 75
of particles, 65
of rigid bodies, 69
stable, 15, 127, 331, 357
Equipotential surfaces, 96
Eriometer, 232
Errors, instrumental, 16
observational, 16
probable, 16
Ether, 179, 202
Evaporation, 152
Exchanges, theory of, 255, 256
Extension, 79

Faraday, 276, 315, 319, 337, 367, 371
Fluid, equilibrium of, 75
motion of, 61, 74
motion of perfect, 74
Fluorescence, 208, 253
Focal lines, 184, 189
Focus, principal, 183, 193
Force, 62
lines and tubes of, 96
moment of, 70
measurement of, 64
Forces, composition of, 65
molecular, 83, 125
range of, 145
Form, 79
Freedom, degrees of, 41, 54, 57
Freezing mixtures, 279
Fresnel, 186, 212, 214, 237, 239, 240, 245
Fresnel's rhomb, 252
Friction, 62

Galvanometer, 369
Gas battery, 341
Gaseous pressure, 149
Geysers, 275
Glaciers, motion of, 273
Graphical method, 17
Gravitation, 8, 62, 85-101

Gravitation, law of, 87
Gravity, centre of, 88
hypotheses in explanation of, 91, 92
Green, 245, 380
Gyration, radius of, 70
Gyrostats, 58, 156

Haidinger's brushes, 249
Hall's effects, 373
Halos, 198
Hamilton, Sir W. R., 201, 246
Harmonic motion, simple, 51, 52
Harton experiment, 90
Heat, conduction of, 282, 287
convection of, 282, 288
latent, 272, 274, 276
mechanical equivalent of, 289
radiant, 253
radiation of, 257-259, 261
specific, 268, 269, 271, 301
total, 276
Helmholtz, Von, 173, 209, 249, 338
Hoarfrost, 277
Hodograph, 48
Hooke, 135, 147, 237
Hooke's Law, 135
Hope's experiment, 264
Huyghens, 186, 234, 235, 237
Hygrometer, 277
Hypothesis, 15
Hysterēsis, 355, 356, 357

Image, 182
real, 183, 195
virtual, 183, 195
Impact, 68
Impenetrability, 79
Impulse, 68
Indicator diagram, 294-297
Inductive capacity, specific, 315
Inertia, 6, 79
centre of, 69, 88
moment of, 70, 71
Isothermals, 24, 295

Joule, 147, 148, 289, 303, 357
Joule's Law, 342

Kepler's Laws, 86
Kerr's effects, 374
Kinetics, 63
Kirchhoff, 203, 204, 256
Kirchhoff's Laws, 336

Least action, 200, 201
confusion, circle of, 184
time, 200, 201
Lenses, 193, 194
achromatic, 197
Le Roux, 207, 327, 333
Le Sage, 91, 147
Leyden jar, 314
Light, absorption of, 202-209
colour of, 180
diffraction of, 224-233
dispersion of, 196, 197, 207, 209
intensity of, 177
interference of, 210-223, 247
nature of, 179, 242
polarisation of, 237-252, 371
propagation of, 177, 186, 225
radiation of, 202-204
reflection of, 181-186, 240
refraction of, 187-200, 240
refraction (double), 234-236, 244, 245, 250
scattering of, 181
speed of, 178
theories of, 179, 185, 186, 199, 200, 375
Lightning rod, 322
Lloyd, 213, 246

Maccullagh, 239, 240, 243, 245
Magnetic axis, 349
force, lines of, 350
induction, 351
intensity, 351
moment, 349
permeability, 352, 354
pole, 349
potential, 350
retentiveness, 353
shell, 362
susceptibility, 352, 354, 356, 357
viscosity, 354, 355
Magnetism, residual, 353
terrestrial, 359
theories of, 360, 365
Magnetometer, 358
Magnets, coercive force in, 353, 356
permanent, 344
temporary, 344
Malleability, 83
Malus, 201, 237, 238, 243
Mass, 7
measurement of, 64

Matter, 4, 5
definitions of, 77
general properties of, 79
heterogeneity of, 142, 145, 146
special properties of, 80, 83
specific properties of, 81
states of, 78
Maxwell, 24, 108, 145, 147, 152, 153, 154, 250, 319, 320, 360, 372, 375, 376
Metallic reflection, 206, 243
Microscope, 195
Mirage, 191
Molecules, 82, 154
size of, 146
Moment, 49
Momentum, 64
Motion, laws of, 63, 64, 65, 68
of centre of inertia, 69
of connected particles, 68
of non-rigid solids, 73
of particles, 64, 65, 66
of rigid body, 56, 57
Motor, electric, 366
Musical intervals, 167

Nebular hypothesis, 93
Neumann, 239, 240, 245
Newton, 63, 67, 68, 87, 88, 92, 186, 223, 237, 259
Newton's rings, 221
Note, 157

Ohm's Law, 336, 338
Osmose, 117

Paramagnetism, 346
Paraselenæ, 198
Parhelia, 198
Path, mean free, 148, 153
Pendulum, 9
ballistic, 68
Peltier effect, 331
Period, 51, 53
Phase, 51
Phosphorescence, 208, 253
Photometer, 177
Piezometer, 113
Polarising prisms, 252
Porosity, 79
Position, 39
Potential, 94
gravitational, 95
Precession, 58
Prévost, 147, 203, 255
Primary cells, 340
Prisms, 192
Projectiles, 47
Pyrheliometer, 260
Pyro-electricity, 232
Pyrometers, 267

Quartz fibres, 89

Radiometer, 153
Rainbows, 198
Rankine, 254, 295
Reflection, total, 188
Refraction, conical, 246
index of, 187, 223
Regnault, 264, 265, 269, 271, 275, 276, 277
Residual discharge, 321
Resonance, 173
Restitution, co-efficient of, 68
Rigidity, 79, 83, 128, 131, 156
flexural, 133
torsional, 131
Ripples, 76
Rotation, 54, 55
Rotational equilibrium, 72
Ruhmkorff coil, 367

Scalar quantity, 40
Schehallion experiment, 90
Secondary cells, 341
Shear, 59
Solar radiation, 260
Solenoid, 365
Solution, 279
Sound, diffraction of, 164
intensity of, 161
interference of, 165, 175
pitch of, 166
propagation of, 158
quality of, 174
reflection of, 162
refraction of, 163
speed of, 159, 160, 172
wave-length of, 157
Spectrometer, 204
Spectrum analysis, 204
diffraction, 233
heat, 258
refraction, 196
Speed, 41
average, 45
mean square, 148
Specific heat of electricity, 332

Statics, 63
Stewart, Balfour, 203, 256, 259, 261
Stokes, 208, 217, 224, 225, 234, 239, 242, 245, 379
Strain, 59, 60
Stress, 68
Surface-tension, 120, 125, 126, 127
Syren, 166

Tait, 153, 278, 284, 330, 333
Telephone, 9
Telescope, astronomical, 195
Temperature, 262
absolute, 293
absolute zero of, 266, 303
critical, 24, 278
measurement of, 267
triple point, 24
Tenacity, 83, 138
Tensor, 42
Theory, 15
mathematical, 15
Thermal capacity, 268
conductivity, 283, 284
Thermodynamic motivity, 298
Thermodynamics, first law of, 289
second law of, 292
Thermo-electric diagram, 330, 333
power, 328
Thermometer, 267
hypsometric, 275
wet and dry bulb, 277
Thermometric conductivity, 284
of crystals, 286
of earth's crust, 285
Thermopile, 328
Thomson, 15, 92, 101, 126, 137, 141, 146, 148, 155, 209, 260, 273, 292, 293
Thomson effect, 332
Tone, 157
combination, 176
partial, 173
Transpiration, 111
Twist, 57

Vector, 40
Velocity, 41, 46
angular, 43, 58
average, 45
Vibrations of air-columns, 172
plates, 170
rods, 168, 169
strings, 171
Viscidity, 115
Viscosity, 79, 83
of gases, 108, 151
of liquids, 115
of solids, 136
Vortex atoms, 141, 156
motion, 61
Vortices, molecular, 254, 372

Water equivalent, 269
Wave-length, 53
motion, 53
Waves, long or solitary, 76
on stretched cords, 73
oscillatory or free, 76
Weight, 64, 79, 85
Wheatstone's bridge, 343
Work, 7, 62

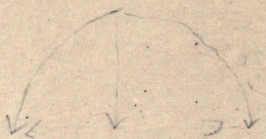
Young, 211, 221, 222, 232, 237
Young's modulus, 130

Zone plates, 230

THE END.

Baillière, Tindall & Cox, King William Street, Strand.

YB 55437



557598

QC
21
P44

UNIVERSITY OF CALIFORNIA LIBRARY